REAL-TIME ESTIMATION OF ARTERIAL PERFORMANCE MEASURES USING A DATA-DRIVEN MICROSCOPIC TRAFFIC SIMULATION TECHNIQUE

A Dissertation
Presented to
The Academic Faculty

by

Dwayne A. Henclewood

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Civil and Environmental Engineering

Georgia Institute of Technology
August 2012

REAL-TIME ESTIMATION OF ARTERIAL PERFORMANCE MEASURES USING A DATA-DRIVEN MICROSCOPIC TRAFFIC SIMUALTION TECHNIQUE

Approved by:

Dr. Michael P. Hunter, Advisor
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Michael D. Meyer
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Randall L. Guensler
School of Civil and Environmental Engineering
Georgia Institute of Technology

Dr. Richard M. Fujimoto
Computational Science and Engineering Division
Georgia Institute of Technology

Dr. Michael O. Rodgers
School of Civil and Environmental Engineering
Georgia Institute of Technology

Date Approved: May 17, 2012

ACKNOWLEDGEMENTS

My tenure at the Georgia Institute of Technology has capped off an amazing journey. Along this journey, I have been helped, guided, inspired and motivated by a host of precious individuals, to whom I will be forever indebted. Their help along the way is as unique as the relationship that we share and their everlasting impact has been experienced through non-verbal 2-second conversations to on-going conversations that are currently guiding and preparing me for the next phase of my journey.

During my time at Georgia Tech, I have had the privilege of being mentored and advised by a number of professors, chief among them Dr. Michael Hunter. Dr. Hunter has played and will continue to play an integral role in my life. His influence has gone beyond that of my transportation career. He has caused me to be even more committed to being an individual that strives for excellence in all of life's arenas, while actively contributing to the success of those who surround me and those that follow. The lessons learned from Dr. Hunter cannot be overstated nor can his contribution to my person be fully expressed. To Dr. Hunter I owe more than I will ever be able to repay. Dr. Hunter, thank you.

To the rest of my doctoral thesis committee, Dr. Randall Guensler, Dr. Michael Rodgers, Dr. Michael Meyer and Dr. Richard Fujimoto, I express my deepest gratitude. Your unwavering support throughout this process has been invaluable. I am truly looking forward to applying all that I have learnt from you as I continue my pursuit of excellence.

To the other members of Georgia Tech's Transportation Systems group, Dr. Adjo Amekudzi, Mrs. Lisa Baxter, Dr. Laurie Garrow, Dr. Angshuman Guin, and Dr. Jorge Laval, thank you for your support and words of encouragement along the way. Without

you, the success that I have experienced throughout my tenure here would have been improbable.

The most valuable aspect of my time here at Georgia Tech is the lifelong relationships and friendships that I have developed over the years. In addition to the relationships with my professors, my classmates and peers have made my time here mind-opening, pleasurable and unbelievably rewarding. Without you all, I have no idea how this chapter of my life would have been written, and more importantly how it would end; for this I thank you. I am unfortunately unable to list all those who contributed to this phase of my life but I do want to say a special thanks to; Shawn Pope, Alek Pochowski, Elise Barrella, Wonho Suh, Narasimham (Lakshmi) Peesapati, and Mshadoni Smith – this has been an amazing journey, let's do it again sometime, preferably without homework.

To my crew that got me here and never left my side – Azizi, Jason, Marsha, Mike, Omar, and Kerry – thanks for being there, constantly inspiring and consistently whispering - "You can do it". Thanks.

In closing, I would like to thank my A-Team - my family. To my ever-praying and supportive parents - Glenford and Pauline, my number one backers, my sisters - Marcia, Venecia, and Monique, my nephew Ryan, who I live to be an inspiration for, THANK YOU. Also thanks to my family, States side, Mark, Mary, Marty and Justin DeFrancisco, thank you for your love and for constantly reminding me that home is really never too far away. To Ms. Jessica DeFrancisco, to my "rock", to my love – I thank you with all my heart for being there for me every time and anytime. It is going to take a life

to express my gratitude to you, and even then, that may not be enough for all that you have done for me, but I hope it will be a start. Jess, thank you.

To all those that I have mentioned and those that I am unable to list your generosity, time, sweat, support, laughs, inspiration, and just putting up with me could not be more appreciated. Thank you.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF SYMBOLS AND ABBREVIATIONS

ax	Average standstill distance
bx_add	Additive part of safety distance
bx_mult	Multiplicative part of safety distance
d	Safety distance
D	Kolmogorov-Smirnov test statistic
dsd_rng	Desire speed range
GDOT	Georgia Department of Transportation
H	Heuristic Form Fit test statistic
ln_chng_dist	Lane change distance
M_f	Final set of model parameter sets
M_n	Complete set of parameter sets
mn_hdwy	Minimum headway
mx_dc_coop	Maximum deceleration for cooperative braking
mx_dc_ow	Maximum deceleration (own)
mx_dc_tr	Maximum deceleration (trailing)
NGSIM	Next Generation Simulation
sf_dist_fac	Safety distance reduction factor
TT	Travel time
U	Mann-Whitney test statistic
v	Mean desired speed

SUMMARY

Traffic congestion is a one hundred billion dollar problem in the US. In 2010, Americans spent approximately five billion additional hours and purchased an estimated two billion gallons of additional gas due to congestion [1]. The cost of congestion has been trending upward over the last few decades, but has experienced slight decreases in recent years as a result of changes in demand and the impacts of congestion reduction strategies.

There has been a fundamental shift in the manner in which congestion issues are addressed. Previously, congested roadways were remedied by increasing roadway capacity. However, acquiring the right of way for roadway expansion has become increasingly difficult and expensive. Thus, there has been a significant push by the government, private industry, and the research community to develop and implement alternate means of alleviating congestion.

A wide variety of advanced technological tools have been developed and implemented throughout the nation's transportation network in an effort to increase its efficiency. Some of these tools include Advanced Traffic Control System (ATCS), Advanced Traffic Management Systems (ATMS), Advanced Traveler information System (ATIS), and Ramp Metering and Managed Lane Strategies. These systems are currently credited with reducing or maintaining freeway congestion levels in light of increasing travel demands.

More specifically, these gains in efficiencies are realized through the work of Traffic Management Centers (TMCs) located throughout the nation. TMCs' efforts to provide accurate and current traffic information to the public are leveraged by and

bolstered through partnerships with federal and private entities. Such entities include the federal government's Traveler Information Telephone Number system (511), Google, NAVTEQ, and INRIX. The information provided by these entities, albeit concentrated on freeway operations, has aided in the increased efficiency of the network. This research effort aims to compliment the currently available real-time freeway information by addressing the lack of real-time arterial traffic information.

The goal of this research effort is to develop a methodology that provides current arterial performance measures, in real-time, to travelers and transportation facility managers by employing point sensor data to drive an online microscopic traffic simulation model. Providing such information to Departments of Transportation and the public will hopefully allow for more informed decisions regarding efficient usage and management of the nation's transportation facilities.

To accomplish the goal of this research, a series of tasks were undertaken; not to only to develop the method to provide real-time estimates of arterial performance measures, but to also explore the feasibility of the implementation of such a method. To begin the method's development, an extensive review of past efforts was conducted. This review revealed that although there have been advancements in providing traffic performance measures in real-time; these advancements have not been equally beneficial to arterials and freeways.

Today, there are a number of available outlets that provide relatively accurate freeway performance measures, including Google and a myriad of commercial GPS-based devices and services. However, for arterial streets, the availability of such information is in its infancy. Currently, estimates of arterial performance measures are

being pursued through the exploration of a number of mathematical techniques that employ vehicular input and other related roadway data. In spite of the successes of these analytical methods, a number of factors have limited their real-world implementation. Five of the more prevalent limitation include: (1) the lack of adequate field implementation, (2) the less than accurate estimates of performance measures, (3) the lack of transferability from one location to another, (4) the dependency on data that may not be readily available from the field, and (5) the lack of evidence to support the functionality of these methods in a real-time framework. This research will learn from the lessons offered by previous efforts to inform the development and implementation of a method that estimates arterial performance measures in real-time.

Building on the strides made by past efforts, the basis of this method uses individual vehicle records, from point sensors, to drive an online microscopic simulation of an arterial. These point sensors are capable of estimating and transmitting individual vehicle information such as vehicle speed, presence and headway. This data is then fed into an online simulation that then interprets and implements the data accordingly into the simulation environment, VISSIM in this case. This simulation environment is also outfitted with the proper roadway geometries and other relevant data, such as signal timing plans, to accurately reflect traffic performance throughout the study area. In addition, this methodology's framework also depicts the ability for this method to extend into the exploration and likely selection of multiple future scenarios. However, given temporal and financial constraints, exploring future traffic states is outside the scope of the current effort.

To evaluate the proposed method's ability to estimate arterial performance measures in real-time, four experiments were conducted. Each experiment built on the success of the previous one, while addressing previously encountered issues. The first experiment was conducted in a test-bed that was comprised of two VISSIM simulation instances. One instance represented the study area, and the other represented the model of the study area. The results from this experiment were encouraging as travel time measures from both instances were approximately equal. Despite this demonstration of feasibility, when both instances were "perfectly" calibrated to each other, it facilitated more agreement between the results than what would have been possible had study area been a field environment.

The second experiment test replaced one of the VISSIM instances with an actual study corridor. A corridor on Georgia Tech's campus was selected and outfitted with temporary detectors. Each detector was responsible for monitoring a detection zone on a corridor's entering link. As a vehicle passes through the detection zone, a message communicating the vehicle's presence and approximate speed was transmitted to a VISSIM instance in the lab. This VISSIM instance was then populated according to the details of the received messages. Although the estimates from this experiment were less accurate than those from the first, these results were still very encouraging. Two of the primary reasons for the reduction in accuracy from experiment 1 to 2 were that: (1) the model was not perfectly calibrated to the field and (2) the traffic signals were not synchronized in the two environments. A third experiment was conducted to address some of these issues while increasing the fidelity with which a model of the study area estimates field measures.

The third experiment was conducted in an on-campus test-bed that was outfitted with a video detection system (VDS). This system monitored several locations throughout the test-bed and transmitted the necessary individual vehicle records to a VISSIM model of the area. Once again, the VISSIM model was appropriately populated according to messages that were received. The simulation results from this experiment demonstrated that a better calibrated model, as well as synchronized signals improves accuracy. The results from the third experiment highlight the accuracy with which the simulated environment can reflect field performance measures in real-time, as well as the viability of the method proposed.

In light of the success of the third experiment, a fourth experiment was designed to explore the potential accuracy with which the simulated environment could estimate field performance measures. The fourth experiment is considered a pseudo-field test. This experiment replaced the field with an origin-destination (OD) matrix that was used to populate a VISSIM model as if vehicles were traversing a corridor and point sensors were transmitting their relevant information. The OD matrix, along with high resolution vehicle trajectory data and signal timing information were obtained from the Federal Government's Next Generation Simulation (NGSIM) program. With such a detailed database representing traffic behavior along a corridor, this pseudo field experiment facilitated the exploration of the proposed method performance when given near to "perfect" data.

It was hypothesized that, with "perfect" data, the simulated environment's estimates of field performance measures ought to be the most accurate set of estimates from among the experiments that were conducted. Although the results from this

experiment were in fact the most accurate set of results, there were still noticeable discrepancies between the simulated estimates of performance measures and those collected in field. After carefully studying the discrepancies between field and simulated performance measures, it became apparent that the VISSIM model needed to be better calibrated to improve the representation of driver behavior along the study corridor. An advance calibration procedure was therefore introduced to further calibrate the VISSIM model of the NGSIM study corridor.

The advanced calibration procedure went beyond the commonplace criterion of calibrating a simulation model to the mean of performance measures. The procedure employed a two-part criteria process. The first part ensured that the simulated estimates of saturation flow where within a reasonable range from estimations in the field. The second part involved an extensive statistical comparison of travel time distributions. After applying this two-part criteria calibration procedure, more accurate estimates were obtained from the simulated environment.

In addition to increasing the accuracy with which the simulated environment estimated field performance measures, this advanced calibration procedure also afforded an opportunity to explore the relationship between calibration parameter values and their ability to produce a calibrated model. This exploration revealed that the values for average standstill distance, additive safety distance, multiplicative safety distance, and desired speed range are seemingly four of the most impactful parameter values when calibrating a VISSIM arterial model. As such, special attention should be paid to these parameters and their values in future calibration exercises.

The results from this calibration process also put forth the case for a real-time calibration algorithm. Previous calibration efforts tend to produce a single model, and therefore a single set of parameter values for a model. The calibrated parameter set is then used to represent traffic conditions at multiple periods throughout a day. Upon analyzing the results from this calibration exercise, different sets of parameter values were able to sufficiently reflect field performance measures for the different time periods that were included in the NGSIM program. Only one set of parameter values was able to produce sufficiently accurate field estimates for all the time periods that were studied. Putting in place a real-time calibration algorithm will enable the proposed methodology to be sufficiently robust in its ability to reflect field performance measures in spite of changes to driver behavior or other dynamic traffic traits that may change over the course of a day.

The results from the aforementioned experiments are encouraging and they demonstrate the feasibility with which the proposed data-driven microscopic simulation method is capable of accurately estimating field performance measures in real-time. Given the success of this method, this research will make a primary contribution to the ongoing effort to increase efficiency when moving people and goods through the transportation system by providing additional traffic information. More specifically, this research will serve as a template for future efforts geared towards real-time simulation for improved (arterial) traffic performance monitoring and operation. The two-part criteria calibration process is intended to inform future calibration efforts as its robust nature and ability to produce well-calibrated models will improve today's predominant calibration methods. The framework in which this methodology was developed also included the

potential to predict future traffic states and performance measures. Although the prediction of future states was not included as a part of this effort, the output from the developed method is intended to be easily incorporated in future prediction efforts. This research, in turn, serves to aid the development of techniques to predict future traffic states.

Despite the success of the method, there are still opportunities to improve the accuracy of the simulated estimates. These opportunities include 1) means of improving vehicular volume accuracy throughout the network, 2) real-time turning movement estimations, 3) signal state transmission, 4) sensitivity to detector loss and inaccuracies and 5) integration and use of probe vehicle data to improve the method's robustness. In addition, devising a means of appropriately accounting for pedestrian behavior, downstream boundary traffic conditions, and a more stream lined calibration process will further speak to the method's ability to estimate performance measures in real-time while speaking to its implementation potential.

CHAPTER 1

INTRODUCTION

1.1 Background and Motivation

Traffic congestion is a one hundred billion dollar problem in the US. In 2010, Americans spent approximately five billion additional hours and purchased an estimated two billion gallons of additional gas due to congestion [1]. The cost of congestion has been trending upwards over the last few decades, but has experienced slight decreases in recent years due to changes in demand and the impacts of congestion reduction strategies, Figure 1.

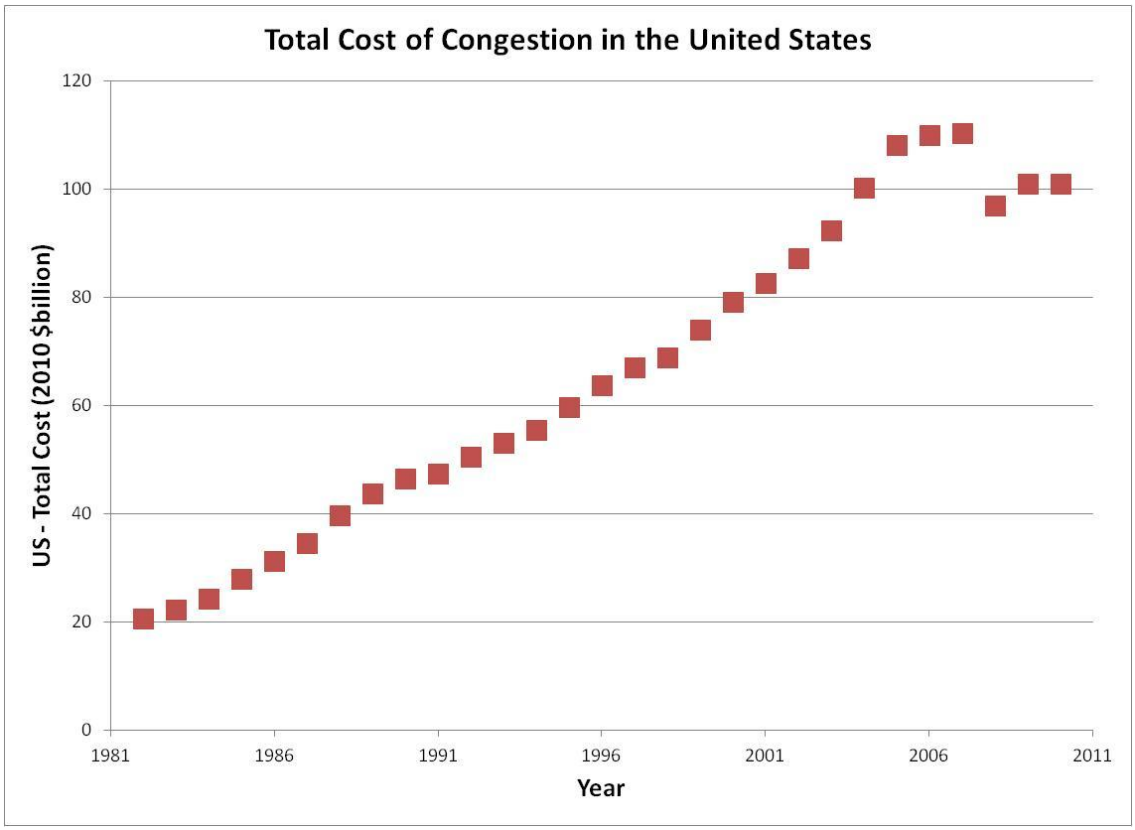


Figure 1 Total Cost of Congestion in the United States [1]

There has been a fundamental shift in the manner in which issues of congestion are addressed. Previously, congested roadways were remedied via an increase in roadway capacity. However, acquiring the right of way for roadway expansion has become increasingly difficult and expensive. Thus, there has been a significant push by the government, private industry, and the research community to develop and implement alternate means of alleviating congestion. This research project explores the feasibility of integrating real-time data streams with an arterial simulation. Such integration is geared towards providing the Departments of Transportation and the public with current estimates of arterial performance measures. This additional information will facilitate increased efficiency in facility utilization by enabling more informed decisions in the use and management of the nation's transportation facilities. To accomplish this, fixed sensors were utilized in the development of an online, data-driven, microscopic traffic simulation tool to determine and provide arterial performance measures. Critical to the success of this tool was the development and application of an advanced calibration procedure that enabled the traffic simulation to estimate field measures more accurately.

1.2 Problem Statement

A wide variety of advanced technological tools have been implemented throughout the nation's transportation network to increase efficiency. Some of these tools include Advance Traffic Control System (ATCS), Advance Traffic Management Systems (ATMS), Advanced Traveler information System (ATIS), and Ramp Metering and Managed Lane Strategies. Currently, these systems are credited with reducing or maintaining freeway congestion levels in light of increasing travel demands. As an example, in Georgia these benefits are primarily observed in the Traffic Management

Center's (TMC) freeway monitoring and quick response when ridding the roadway of any obstacles that may reduce freeway service levels. There have been a number of efforts to leverage the work done by TMCs to provide travelers with more current traffic information such as Georgia 511 / Navigator [2]. In addition, private efforts and partnerships with companies such as Google, NAVTEQ, and INRIX have made the TMC's information more accessible to travelers, aiding their traveler decisions [3–6]. This effort aims to compliment real-time freeway information by addressing the lack of available real-time arterial performance measures.

In comparison to the vast investments in equipping freeways with advanced technology to improve mobility, widespread outfitting of arterials with similar technologies in its early stages. Successful ITS arterial deployments include both advance and adaptive traffic signal control systems and various surveillance efforts. The benefits of these limited deployments range from a reduction in the number of stops along an arterial segment to increases in traveler satisfaction [2]. Also, more recently, real-time traffic information providers have been supplying information regarding traffic condition along arterials, the accuracy of which is still being improved.

Table 1 and Table 2 highlight the current differences in the ITS related benefits as experienced by travelers on arterials and freeways, respectively. The difference in disseminating ITS information to the traveling public is noticeable. Table 2, referring to freeways, demonstrates that presenting traffic information to the travelling public has positive impacts on safety, mobility, and customer satisfaction. From Table 1 (referring to arterials) one notices that there is insufficient data to support a conclusion regarding the benefits of disseminating traffic information to the travelling public along arterials.

One of the reasons for this lack of conclusion is that the information disseminated is very limited. Of the arterial streets network in the nation's largest 108 metropolitan areas, arterial traffic information is only available for approximately two percent of the network miles [2]. This effort seeks to address this lack of available real-time arterial traffic information and aid in the realization of all possible benefits that may be experienced by potential travelers, drivers and transportation facility managers.

Table 1 ITS Arterial Management Benefits Summary [2]

	Safety	Mobility	Efficiency	Productivity	Energy and Environment	Customer Satisfaction
Surveillance						
Traffic Control	+	●	●		●	
Lane Management						
Parking Management		●	+			●
Information Dissemination						
Enforcement		●				●

●	Substantial positive impacts	+	Positive impacts
○	Negligible impacts	*	Mixed results
×	Negative impacts	**blank**	Not enough data

Table 2 ITS Freeway Management Benefits Summary [2]

	Safety	Mobility	Efficiency	Productivity	Energy and Environment	Customer Satisfaction
Surveillance	**Enabling Technology**					
Ramp Control	+	●	●			●
Lane Management	+					
Special Event / Transportation Management						
Information Dissemination	+	+				+
Enforcement	●					●

●	Substantial positive impacts	+	Positive impacts
○	Negligible impacts	*	Mixed results
×	Negative impacts	**blank**	Not enough data

1.3 Research Objective

As stated, the overall objective of this research is to determine the feasibility of integrating real-time data with an arterial simulation to estimate performance measures in real-time and provide such information to facility managers and travelers. This objective was accomplished by under-taking the following tasks:

1. Describe the current state of practice concerning the estimation of real-time arterial performance measures.
2. Develop a federated (integrated) simulation test-bed for testing procedures and algorithms.
3. Determine the feasibility of integrating point sensor data with simulation to create a data-driven, on-line simulation tool.

4. Develop procedures and algorithms to calibrate an on-line simulation tool that estimates travel time and other performance measures in real-time.
5. Explore any potential improvements in travel time estimation resulting from sensors placed in atypical locations, such as immediately downstream of an intersection.
6. Field-test the data-driven, on-line arterial simulation tool on a target corridor.
7. Develop a robust calibration procedure to increase the accuracy of simulation output, while informing the future implementation of such a procedure for real-time calibration.

These objectives are closely aligned with those established by the Transportation Operations and Safety Group at the Georgia Institute of Technology to explore the feasibility of integrating real-time data into an arterial simulation on behalf of the Georgia Department of Transportation (GDOT). This research effort was partly funded by GDOT as the Department was interested in the long-term goal of utilizing real-time arterial data as input for improved control strategies. GDOT was also interested in providing real-time arterial operational knowledge to promote more a efficient and effective use of the arterial system by the traveling public. As such, the results and lessons learned through this research effort will be shared with GDOT as they look to improve the services offered to Georgia's traveling community.

1.4 Dissertation Organization

The organization of the remainder of this dissertation is as follows. Chapter 2, Literature Review, provides a comprehensive review of previous works that made strides

towards estimating arterial performance measures and highlights how this effort will build upon those previous efforts. Chapter 3, Methodology, presents the methodology that has been developed to achieve the objectives of this research project. Chapter 4, Experimentation and Evaluation, details the execution and results of a number of field experiments that were used to validate the developed methodology. Chapter 5, Model Calibration, presents an advanced calibration procedure that was developed and used to calibrate a VISSIM simulation model. Chapter 6, Future Research, offers readers a few tasks that will be tackled in the future to increase the robustness of this method. Finally, Chapter 7, Conclusion and Contribution, concludes this dissertation by highlighting the manner in which the objectives of this task were accomplished and the contributions that this research will make to the transportation community.

CHAPTER 2

LITERATURE REVIEW

The following chapter provides a comprehensive review of previous efforts in estimating and predicting performance measures along signalized arterial streets. A number of estimation models, along with their successes and contributions to the field of estimating performance measures will be presented. Models that have been developed to predict performances measures along arterial will also be reviewed.

2.1 Estimating Arterial Performance Measures

Estimating performances measures along arterials is often more challenging than for freeways. The primary reason for this is that freeways are controlled access facilities with limited mainline traffic control (i.e. no signals, stop signs, etc.) while arterials are often uncontrolled (or limited control) access facilities. That is, vehicles may turn on and off the facility at multiple locations, interaction with potentially numerous crossing arterials may be significant, and control devices (e.g. traffic signals) can significantly influence vehicle movements. As a result of such interruptions volume and speed are extremely varied. Given the large variations in speed and volume along arterials, the ability to determine performance measures can be dependent on significant data needs and more advance mathematical techniques than those employed to extract performance measures from freeways.

2.1.1 The Early Models

In 1977 P.G. Gipps provided one of the earliest models for estimating performance measures along arterials. Gipps developed a regression model that employed

occupancy measurements and vehicle arrival times, from loop detectors, to estimate arterial link travel times. The model was then validated using simulated data. Despite several model adjustments and the relative success of the model, Gipps noted that in order to improve the accuracy of his model, incorporating other parameters such as signal timing plans, number of lanes, and link length was needed. In building on the accomplishments of Gipps' 1977 model, a number of researchers used his model as a foundation for their own model to improve the estimation of arterial performance measures. Gault and Taylor sought to improve Gipps' 1977 model by calibrating it to a two lane roadway and eliminating a few of the parameters that they deem to have minimal impact on the relevant performances measures. [7–9]

A review of a number of the earlier works, including the two previously mentioned models, was conducted by Sisiopiku and Pouphail [8]. The limitations presented ranged from the lack real-world validation results to use of assumptions that may prevent the respective model from being implemented in the real-world. Table 3 presents a number of early models and their associated limitations and validations results [8].

Table 3 Limitation and Validation Results for the "Early" Models [8]

Model	Limitations	Validation
Gipps	-Lack of empirical validation -Signal settings/geometry not considered -Correlation of the parameters exists	With simulated data only; MSE[a]=0-15%
Gault et al.[b]	-Underestimates travel time occ.>50% -Lack empirical validation	With simulated data only; Within 10% of the mean
Gault	-Bounded (occ. should be ≤70% -Not appropriate for oversaturation	With video tape data; Within 10% (rarely up to 50%)
Abours	-Signal setting are ignored -Formulation not reported	With floating car data; RMSE[d] =13%
Luk et al. **Usami**	-Requirement of stop-line detectors -Applicable for oversaturation only	Not reported With simulation & field data RMSE = 10-19%
Luk	-Flow conservation assumption -More suitable for freeway environment -Requirement of stop-line detectors	With wheelbase data With 10% of the mean
Takaba	-Linearity assumption between travel time & flow in congestion -Neglect of dependency between links	Error ratio =12-24%

[a]MSE: Mean Square Error
[b]Arrival Type Model
[c]Occupancy Model
[d]RMSE: Relative Mean Square Error

Zhang and Kwon also presented an overview of a few of the earlier models that were used to estimate arterial performance measures. In this report, the authors grouped the models being studied into five (5) categories. These model categories are regression, dynamic input-output, pattern matching, sandglass, and models developed by the Bureau of Public Roads (BPR). The following sections will briefly highlight the characteristics of these categories as well as their limitation as it pertains to estimating performance measures along arterial. [10]

2.1.1.1 Regression Travel Time Estimation Models

Regression models attempt to use data that is currently available via today's surveillance and control systems. These models are capable of accounting for the various different factors that may affect arterial travel time, however, the models often become location specific and difficult to transfer to other arterials. One of the main similarities among the different regression models is required input data. The input data needed for these may include time registered by a vehicle on a loop detector, occupancy (derived from a loop detectors), offsets, and other signal parameters. Despite the applicability of these models, their estimation of travel times, when compared to those from the field, are often less than satisfactory and therefore in need of further improvements. [7], [9], [10]

2.1.1.2 Dynamic Input-Output Link Travel Time Models

Generally these models use input-output traffic flow relationships, measured at upstream and downstream detectors, along with assumptions describing the change in flow characteristics between the detectors. This class of models is able to estimate both link and route travel times using minimal site specific data. However, a disadvantage of these models can be an inability to predict travel times (as opposed to estimating current travel times) and they require a greater data sampling rate than what is currently available with the use of today surveillance equipment. [10], [11]

2.1.1.3 Pattern Matching Models

In pattern matching the upstream loop detectors record a sequence of voltage signatures from various vehicle types. This sequence of voltage signatures is then compared to those collected from a downstream loop detector. The time between upstream and downstream matching sequences is the average travel time. This approach

can also be used to estimate other performance measures such as traffic density and space mean speed. A challenge to pattern matching approaches is that they often require a data sampling rate and accuracy that is higher than that which is obtainable from today's field detectors. [10], [12] A more recent example of a technology that has demonstrated success using pattern matching is the wireless traffic detection and integrated traffic data systems offered by Sensys Networks [13].

2.1.1.4 Sandglass Link Travel Time Models

These model use the concept that travel-time can be estimated as the sum of time spent on two segments of a link – a congested segment and an uncongested segment. On the congested segment of the link there is no inflow of vehicles from external sources nor is there outflow to other roads, thus travel times are essentially deterministic queuing delays. For the uncongested segment travel time is determined by using a constant speed relationship. One with a challenge of these models is that the required input is queue lengths which may only be indirectly obtained from the field data. Therefore any error in estimating queue lengths from the loop detectors will be propagated throughout travel time estimation. Furthermore, the accuracy of these models is unsatisfactory especially for dynamic short-term traffic management applications. [10], [14]

2.1.1.5 Bureau of Public Roads (BPR) Models

The models developed by the BPR to estimate performance measure along arterials have primarily been used in transportation planning and intersection studies. Like sandglass models, travel-times are computed as the sum of time spent on two segments in a link, the free-flow travel time and intersection delay. The input required for

these models is traffic volume data which is obtained directly from loop and video detectors. However, despite the anticipated accuracy of these model, when tested the result tended to be unsatisfactory. [10], [15]

2.1.2 Developments in Estimation Models

Building on the successes and lessons learned from earlier models, a number of recent efforts have been devoted to addressing the limitations and refocusing the assumptions of earlier models. One of the first significant attempts to build on earlier models was presented by H. M. Zhang in 1998. Zhang developed the Link-Journey-Speed (LJS) model which estimates the speed, and subsequently the travel time, along signalized arterials. The LJS model combines the speeds estimated from the roadway's critical volume to capacity ratio and the one calculated from the volume and occupancy measurements from loop detectors. The model has been demonstrated to work well in under capacity conditions although may break down under congestion conditions - particularly when the built up queues are not long enough to be detected by upstream detectors. [16]

In 2007, Liu and Ma presented a time-dependent model to estimate travel time along arterials. In this paper the authors developed a model that used loop detector and signal status data to calculate travel time along an arterial corridor. When calculating travel time the model decomposes travel time into two components; free flow travel time and intersection delay. Although the presented model estimates travel time along arterials fairly accurately its validation was completed in a simulated environment. Additionally, given that the model greatly relies on loop detector and signal status data, a real-world implementation of this model may be met with a number of challenges relating to data

accuracy and transmitting the data from the field to a remote location to be implemented in the model. [17]

Wang and Hobeika present a modified HCM2000 model to estimate travel time along arterials. Similar to previous models, this model estimates travel times as a sum of free flow link travel time and delay experienced at an intersection. Essential to this model is the speed and volume data collected by upstream loop detectors. Based on these data free flow travel time is a simple calculation involving travel speed and link length while intersection delay is calculated by grouping vehicles together and using the relationship between average intersection delay and number of vehicles per cycle length as a well as average intersection delay and the delay of the first vehicle in a group of vehicles. The proposed model was validated using average intersection delay from a single intersection in the field and delay computed using the HCM 2000 method. Despite the accuracy with which this model estimates intersection delay for a single intersection along an arterial, the authors acknowledge that extending the model to involve a number of intersections will further demonstrate the feasibility of employing this model to estimate arterial travel time. Additionally, a potential limitation of this model is that it requires upstream loop detectors, which are not often times available in the real-world. [18]

2.1.3 Automatic Vehicle Location and Identification Estimation Methods

As technological advancements have been made in the fields of global position systems and various vehicle identification technologies, a number of researchers have employed the use of such technologies to better estimate transportation network performance measures, particularly travel time. Although the usage of these technologies has been largely geared toward freeway implementation there are a number of efforts that are

aimed at extracting performance measures along arterial streets. Dailey and Cathey, in 2002 developed a estimation methodology that used transit vehicles that were equipped with advance vehicle location (AVL) technology, with the aid of Kalman filtering to estimate speeds and travel times along freeway and principal arterials [19]. Li and McDonald in 2002 presented a link travel time estimation model that used GPS data from a single probe vehicle. This model uses the time-speed profile of the probe vehicle to produce a maximum continuous acceleration and an average speed value to be inputted into fuzzy sets. Once these values enter the sets they will be analyzed with historical traffic data to derive travel time along the link being studied. Despite the promising results from this research effort the model's use of a single (or a few probe vehicles) to represent the traffic in its entirety along a particular arterial may provide erroneous data as a particular driver's behavior may not be representative of the traffic's current condition. Furthermore, building fuzzy sets of driving patterns for a large arterial network may be a tedious and labor intensive process. [20]

Choi and Chung in 2003 presented an algorithm the fused data from GPS equipped vehicles and loop detectors to estimate link travel times along arterials. This algorithm also employed the use of a voting technique, fuzzy regression, and Bayesian pooling to aid in the estimation of arterial travel time. The base of this proposed algorithm is a double fusion data process while incorporating the historical traffic data of the link being studied to estimate link travel time. The results from the model indicate that this algorithm does accurately estimated the travel time for the arterial links understudy. However, possible limitations include lack of feasibility in, near-term, real-world implementation of the algorithm given its dependency on GPS and transmitted

loop detector data. Also the authors indicate that further tests need to be done to analyze how the algorithm will perform under different traffic conditions. [21]

In 2009 Pu et al. [22] presented key limitations associated with AVL technologies to estimate arterial performance measures in real-time. To address some of these limitations, the authors developed a framework that employs historic bus and car speeds, and streaming AVL bus speeds to estimate bus and car speeds, and travel times in real-time. Central to this framework is the joint relationship between bus and car speeds which has been formulated through the use of historic car and bus speeds. Despite the method's promising results, accurate estimates are dependent on stream AVL bus data which is not always available. Also, the authors highlighted the need for further studies, before full scale implementations, to evaluate the frameworks performance under changing traffic conditions. [22]

Lucas et al. (2004) [23] presented three noteworthy limitations when using GPS and other forms of vehicle identification technologies. In part, these limitations are associated with the offsite processing of vehicle identification data which hinders real-time implementation of such methods, the cost associated with additional equipment and infrastructure investments and privacy concerns of drivers as they traverse to transportation network. To address these limitations, the authors presented an estimation methodology that only relies on vehicle platoon information from loop detectors. Although promising a disadvantage of this method is that it requires streaming detector data which is a limitation of today's traffic controller. [23]

All of the above efforts attempt to estimate performance measures in real-time. However, this goal has been achieved with varying levels of success and accompanied by

different sets of limitations. Some of these works present an entire method to extract real-time performance measures, albeit with limited success during full scale field implementations. And others are more geared towards improving a particular component of a real-time performance estimation framework and not necessarily developing a complete methodology.

2.1.4 Statistical Models

There is a large body of work of statistical estimation models that are aimed at approximating performance measures along arterials. In this category of models traffic data such as vehicle speed, occupancy, headway, traffic flow volume, etc., are used as input variables for equations or models that output performance measures such as travel time [17]. These models may be divided into sub-categories such as classical statistical models and more complex statistical models. Classical statistical models refer to models that use traditional estimation techniques such as linear, non-linear, and Bayesian techniques to estimate arterial performance measures. The more complex model refers to model that employ techniques such fuzzy logic, neural networks, etc. or any combination of these techniques.

In terms of examples of classical statistics models Turner et al. [24] presented a series of linear expressions to estimate speeds along arterials and subsequently travel time, Zhang [16] presented a non-linear model that combines two speed estimates to calculate arterial travel time and Park and Lee [25] used a simple Bayesian estimator as the basis of a model to estimate arterial link travel speed. As for more complicated models Park and Lee [25] paired a simple Bayesian estimator with an expanded neural network to estimate link travel speeds along arterials, Cheu et al. [26] uses a multi-layer

feed-forward neural network with back propagation training to fuse various data streams to estimate arterial speed, Palacharla and Nelson [27] employed the use of fuzzy logic and neural networks to dynamically estimate arterial travel time and Robinson and Polak [28] considered a k – Nearest Neighbor methodology to determine arterial travel time using loop detector data.

Some of the limiting factors of these models include that they can be site specific and must be recalibrated for different locations and a number of these have only been evaluated under simulated conditions. In addition many of the statistical models require large field data sets not only for the purposes of statistical significance but also for some of the learning algorithms to have more training before estimating arterial performance measures. [17], [18]

2.1.5 Real-Time/Online Estimation Models

Skabardonis and Geroliminis (2005 and modified 2008) proposed an analytical model to estimate travel times along arterial streets in real-time. This model utilizes data that can be had from loop detectors such as, flow and occupancy, and pairs it with signal timing data such as, cycle length, green time, and offset. Kinematic wave theory was then used as the base of this model as it was able to represent the spatial and temporal features on queues formed at signalized intersections. Similar to previous models the travel time on an arterial link is calculate as the sum of the link free flow travel time and the delay experienced at the intersection. In this model the delay incurred at an intersection is equal to the summation of the three forms of delay, the approach delay, queue delay, and delay due to oversaturation. In light of this model's ability to estimate travel time with relatively high accuracy it was validated in a simulated environment and also with limited

field data. However, field data trials where offline, not utilizing a real-time data stream. [29], [30]

Tsekeris and Skabardonis (2004) examined five analytical models that have been primarily develop for use in real-time estimation of performance measures along arterials. These five models are the Spot Speed (SSM), BPR-Based, Uniform Delay-Based (UDM), Overflow Delay-Based (ODM), and the Generalized Delay-Based (GDM) models. The evaluation of these models' ability to estimate performance measures in real-time was conducted in a simulated environment. In their simulated environment they found that to fully evaluate the robustness and accuracy of these methods, aggregated travel times, at the network level, and available signal timing information should be taken into consideration. In general, the GDM and ODM were the most promising approaches to estimate total average travel times – at the network level. While the other models provided better estimates individual link travel times. The GDM and ODM were also capable of improved network-wide travel time estimates and greater output robustness when there are discrepancies between field and simulate signal timings. However, it is not known how these simulated findings would translate to a field implementation. [31]

In 2009 Kwong et al. [32] presented a scheme for estimating the distribution of travel time on an arterial link. This scheme employed the use of wireless sensors to acquire the magnetic signature of each vehicle. An upstream signature is matched anonymously with the signature from a downstream sensor to estimate the travel time of a particular vehicle. The authors also state that other performance measures such as link volume, delay, and queue length can be determined from this methodology as distributions. The means of extracting performance measures from a vehicle's upstream

and downstream magnetic signature is a statistical model of signature distance that requires no additional detector data, such occupancy, or infrastructure data such as signal timing plans. In light of the preliminary success of this model, there is a need for further field evaluations as the current evaluation procedure was done on a simple network. In addition, ground truth verification of determined performance metrics is needed. [32]

Lucas, Mirchandani, and Verma in 2004, [23], presented a methodology to extract travel arterial time information without the need to indentify individual vehicles. Their methodology identified vehicle platoons as they traversed the transportation network. The platoons are identified with the use of loop detectors placed along the arterial corridor being studied. The results presented by the authors are encouraging however based on testing in a simulation environment only. As previously mentioned, Zhang and Kwon [10] highlights that such techniques often encounter difficulties when trying to estimate performance measures in real-time as the sampling rate needed for platoon matching is often not available in the field.

A preliminary study was untaken in Melbourne, Australia to investigate the feasibility of extracting arterial travel time measures in real-time. The study was conducted along a small signalized arterial corridor controlled by SCATS (Sydney Coordinated Adaptive Traffic System). In this approach, SCATS datasets, aggregated in 60 second bins, were used in conjunction with historical travel time data from VicRoads to provide estimates of real-time travel time. A drawback of this approach is that to obtain estimated travel time data the given signal system (SCATS) must also be used. In addition additional detectors for successful field implementation may be required. [33]

A large scale attempt to extract arterial performance measures in real-time was presented by Whale [34]. In this paper, the authors presented a methodology that employed the use of a cellular automaton microscopic traffic simulation software and approximately 750 inductive loop detectors located throughout the study area, Duisburg, Germany, to estimate roadway performance. In essence, this methodology acquires traffic information, namely vehicle counts, from each of the approximately 750 detectors at a resolution of 60 seconds. The data used as input to the cellular automaton traffic simulation model. Upon receiving the data and performing the necessary data processes, the load on each link is then presented to the consumers of this information. Limitations of the approach include: a) the use of a cellular automaton traffic model which has a few deficiencies in representing traffic and driver behavior on a microscopic scale; b) a lack of flexibility in the resolution at which traffic data are sent and processed; and c) the vehicle load along a particular link is the only performance measure being delivered to the consumers of the results this effort. [34]

In another effort a team of researchers from the University of Minnesota developed the SMART-SIGNAL system (Systematic Monitoring of Arterial Road Traffic and Signals). This system is a data collection and performance estimation tool for arterial streets. Integral to the functionality of this system is the collection of high-resolution event based traffic data from an arterial. The primary data sources for the system are signal controller cabinets that are located throughout the arterial being studied. From these cabinets event based data such as vehicle actuations and signal phase changes are collected, archived, and processed. This rich dataset is then archived and processed to determine a variety of performance measures. Estimates of performance measures

include travel time, queue length, and number of stops, under a variety of conditions. The field implantation of this system indicates that it is capable of producing accurate performance estimates in real-time. One challenge of this approach is the requirement to gain access to a signal cabinet to extract the event-based signal data. In addition, this system is more feasible for a corridor which is controlled by a network of controllers with one being a master. Where a master cabinet is not present real-time data acquisition becomes a more significant challenge particularly given the resolution required by this methodology. [35]

From the above sections one realizes that a number of advancements have been made in the field of performance measure estimation along arterials although significant limitations still exist. It is also noted that while many of these above efforts discussed their finding and underlying algorithms they did not present significant information regarding data transmission methods or requirements, the impact of lost data or erroneous detections, required detector data filters, or other implementation issues. Despite the successes of the state-of-the-art methodologies and systems, there are few limitations that this research effort is looking to address while building on the capabilities of these earlier works.

2.1.6 Available Real-Time Traffic Information Services

Currently, there are number of providers that offer traffic information in real-time. A few of the major participants in this arena include Google [4], Traffic.Com (NAVTEQ) [5], INRIX [36], Total Traffic Network (TTN) of Clear Channel Radio [37], and SpeedInfo [38]. Although this short list highlights individual organizations that are currently providing information regarding traffic performance, it is noted that a number

of these and similar organization offer these services in collaboration with similar organizations.

The primary means by which these service providers obtain data to estimate real-time traffic performance measures is through infield sensors and GPS enabled devices. For instance, Google relies on individuals that have their GPS based, mobile Google Maps smart phone application enabled. Google aggregates these individuals' data to estimate the current state of traffic, primarily on arterial streets [4]. As for freeway data, Google as well as other traffic service providers also rely on point sensor data often provided by regional and local transportation agencies, such as departments of transportation. Traffic.com, an affiliate of NAVTEQ, acquires its data from its own network of digital traffic sensors, commercial and government partners, and their own traffic operations centers [5]. SpeedInfo uses its solar powered, DVSS-100 Doppler radar Speed Sensor system which measures the speed of vehicles on both sides of the highway [38]. In addition to some of the previously mentioned data sources INRIX gathers information from GPS enabled commercial vehicle fleets to estimate traffic performance [36]. TTN employs information from Airborne/Mobile Spotter Vehicles, Digital Scanners that cover many local emergency services, Police Callouts, and Traffic "Tip Lines" [37].

Accuracy of the traffic information being offered by these service providers is highly dependent of the facility type and acceptable confidence band for the particular consumer's application. The freeway performance accuracy is commonly higher than that for surface streets. This is primarily due to limited access nature of freeways and more uninterrupted flow characteristics. These attributes of a freeway facility lends itself to accurate performance measures being extracted, particularly on the macroscopic scale,

with a fairly narrow confidence band. As for surface streets, both vehicle speeds and volume are highly variable due to intersections (signalized and unsignalized) and frequent, uncontrolled access points. In addition, to date, only macroscopic level information is available for both freeway and arterial facilities, representing roadway segments instead of individual vehicle performance.

CHAPTER 3

METHODOLOGY

The methodology employed utilizes point sensor traffic data to drive a microscopic traffic simulation in real-time. Detector data were transmitted and used as input to a simulation model of the area being studied. Arterial performance measures are then estimated from the real-time simulation. In describing the methodology this section first presents the conceptual framework for the effort followed by the current implementation status. In the current research effort the microscopic simulation package VISSIM is utilized.

3.1 Conceptual Framework

Figure 2 illustrates the conceptual framework for developing a real-time, online, data-driven simulation tool. The first step in the process is to obtain real-time traffic related data from the network's roadway detectors. These data are then processed by the data processing server. Next, the current traffic state is estimated by streaming the processed detector data into a calibrated simulation model of the area being studied. Once the traffic's current state is captured in the simulated environment, the model may be used to predict near-term future traffic conditions. For example, instances of the traffic's current state may be generated and run faster than real-time to provide a series of possible future traffic states. From these future states, a probable future state may then estimated. The current research effort focuses on the use of real-time data to estimate the current traffic state however future research efforts will seek to extend the current estimation platform for use in near-term traffic prediction.

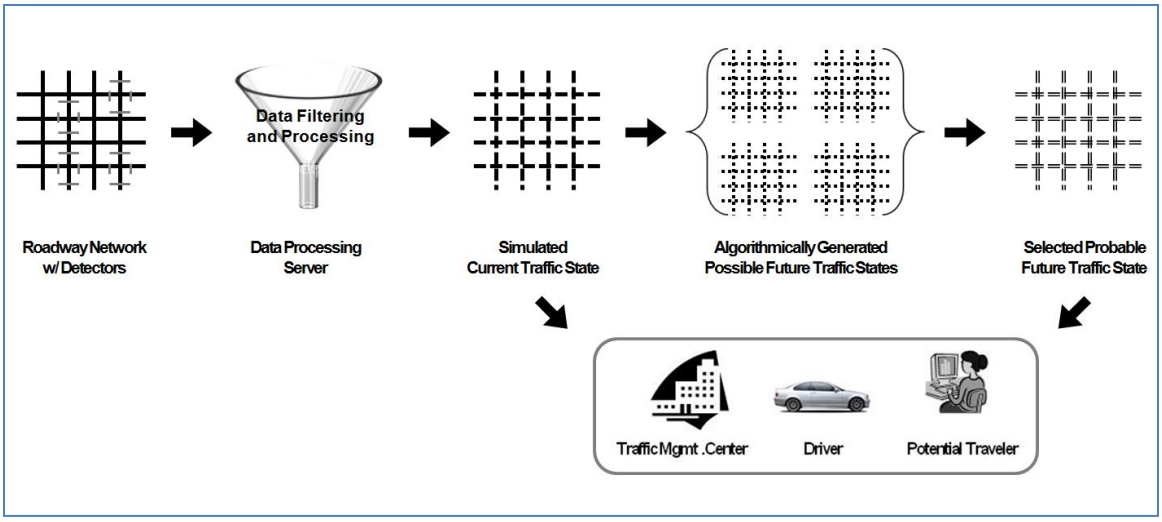


Figure 2 Conceptual Framework for Proposed Methodology

3.1.1 The Network and Detectors

As stated, the goal of this project is to deliver arterial performance measures in real-time using an online data driven microscopic traffic simulation. This research assumes an arterial network where point sensor (i.e. loop detectors, video detection, etc.) detection equipment is available, or may be deployed, capable of transmitting detection data in real-time. It is noted that while real-time transmission capabilities are not commonly utilized such technology exists and is being increasingly adapted. It is further assumed that the detector location is known and may be mapped to the simulation environment. Minimum required data streamed from the detector include individual vehicle actuations and speed. Other traffic related data such as occupancy, headway, and volume may also be available however is not required for the current research effort.

3.1.2 Data Processing and Communication

The communication infrastructure to implement the real-time simulation has three primary tasks: 1) manage the transmission of traffic data between the point sensors and the data processing unit, 2) facilitate the communication between the data processing unit and the simulation, and 3) broadcast the current and most probable future traffic states.

For the current implementation in the first task the data that is sent from the point sensors is processed by a central data processing unit to facilitate implementation of the data into the simulated environment. The data processing unit reads the data from the detector technology and converts this data into the appropriate message format for transmission to the simulation model.

The second task facilitates the passing of information between the processing unit and the traffic simulation. Given the specific requirements for data transmission, processing, and sharing with simulation instance(s), a customized communication tool is employed. This tool is referred to as the Transportation Runtime Infrastructure (TRTI). TRTI is a High Level Architecture (HLA) inspired communication framework that manages the passing of information between clients (i.e. simulations, data processing unit, etc.). For detailed TRTI development, functionality, and implementation information please see Henclewood et.al (2012) [39].

The third task broadcasts the current and estimate future states for use in traveler information systems or in traffic control optimization. A web-based application for presenting the information has been developed. For transportation facility managers, it is also envisioned that in addition to the web-based application they will have access to the raw data. This will allow for the use of model outputs in systems capable of adjusting

traffic system parameters in real-time, such as signal timings, allowing for increased traffic control system responsive.

3.1.3 The Simulated Environment

VISSIM, distributed by PTV, is a high resolution simulation program that is capable of modeling multi-modal traffic flow. VISSIM also has the capability to visually represent traffic. VISSIM also provides a COM (Component Object Model) interface which allows VISSIM to be automated by other applications. The COM interface also provides users access to VISSIM objects, so that they may be created, manipulated, or deleted. For additional information regarding VISSIM and it VISSIM COM interface see [40] and [41].

It is noted that one of the most critical aspects of this research effort is the need to have a well calibrated simulation model of the area being studied. Chapter 5 documents the calibration effort undertaken as part of this research. Current calibration efforts are focused on *a priori* calibration of the model parameters (i.e. vehicle acceleration, look ahead, safety distance, etc.). Future research will explore real-time calibration of VISSIM model parameters. However, a well calibrated base model will remain critical as it is anticipated that the real-time calibration provisions will work best where only small adjustments to VISSIM parameters are required.

3.1.4 Test Bed

Video cameras were selected as the point sensors to be used for this test bed developed as part of this research project. Ten video cameras have been installed in a test bed located next to the Georgia Institute of Technology campus. A Video Detection

System (VDS) was selected as the accompanying hardware and software, facilitated the real-time transmission of event-based traffic data to a remote location. In addition the video detection system is capable of extracting a significant portion of available roadway data. Currently, the ten cameras that have been installed transmit their video via fiber optic cable to the data processing unit. This unit then processes the videos and sends all the relevant traffic data via wired or wireless connection to a client personal computer. This client then parses the data stream and inputs it accordingly into a VISSIM model of Georgia Tech's campus. Figure 3 presents the test bed's location, camera positions and their respective views.

Figure 3 Test Bed Location and Camera Layout and Coverage [4]

3.2 Summary

In the following chapters the above conceptual framework will be expand. This will be in the context of a series of method implementations, ranging from lab implementations to full field tests. These implementation presentations will then be followed by discussions on related research items explored as part of this effort including the use of next to perfect data and model calibration.

CHAPTER 4

EXEPERIMENTATION AND EVALUATION

Four experiments were conducted to determine the feasibility of the proposed methodology. The first of was a proof of concept experiment which was conducted in a simulated environment. The second and third were field experiments, with the primary difference being the use of temporary versus permanent detectors. The fourth experiment was a pseudo field experiment that employed trajectory and origin-destination data to approximate streaming detector data. The following presents the details and results for each experiment.

4.1 Experiment #1: Simulated Environment - Proof of Concept

The proof of concept seeks to provide insight into the feasibility of the proposed real-time simulation framework. This experiment uses two VISSIM simulation instances. One instance represents the "real-world" the other attempts to replicate the "real-world" simulation in real-time (referred to as the "modeled-world"). The inputs to the real-world model include traffic volumes over a 4-hour period (reflective of a peak period), signal timing data, and vehicle turning movements. The modeled-world simulation has the same roadway configuration, signal timing data, and historical turning movement percentages. The modeled-world simulation is not given any vehicular volumes as part of the input files. Instead, as will be discussed, vehicles are generated according to the data obtained from the detectors in the real-world simulation instance.

This initial experiment explores the feasibility of approximating traffic conditions of the real-world simulation in the modeled-world simulation. To determine how well the modeled-world replicates the real-world travel time and delay over representative paths,

and queue lengths at the approaches of the various intersections, are collected and compared.

4.1.1 Experimental Design

A three intersection arterial was created using VISSIM, with each intersection under two-phase semi-actuated signal control. Each roadway is a two-way arterial, with one lane in each direction. In the real-world simulation a loop detector is placed 100 feet from the upstream end of each entrance link, for a total of eight boundary loop detectors. These detectors are responsible for capturing the presence and speed of a vehicle as it enters the network. In both the real-world and modeled-world there are six additional detectors, one on each intersection cross street approach. These detectors are used to implement semi-actuated traffic signal control. No data are currently passed from these detectors in the real-world simulation to the modeled-world simulation. Both models simulate a four hour time period during which the maximum network volume reached is approximately 1200 vehicles/hour and a minimum of approximately 550 vehicles/hour.

A framework in C++ was developed to implement the system shown in Figure 4. In this framework VISSIM COM is utilized to provide a direct means of interacting with a simulation during runtime. To establish communication between the two simulation models a unidirectional named pipe is created. A pipe is a specific section of memory that is used for the purposes of communicating between a server and one or more clients. When using pipes the pipe-server is the process that creates the pipe and the pipe-client is the process that connects to the created pipe [42]. In the named pipe that was created the real-world simulation model served as pipe-server and was able to write to the pipe. The pipe-client was the modeled-world simulation and was able to read from the pipe. While

pipes are capable of two-way communication for the purposes of this experiment a unidirectional pipe was sufficient. Subsequent experiments replace pipes with the TRTI.

Each of the eight detectors that are placed at the edge of the real-world simulation network are polled for vehicle speed, location, and lane data once every simulation second. In this example, given the fixed detector locations, a detector ID would be sufficient in place of the location and lane data, however, passing location and lane data was undertaken to allow for more robust data streams in future experimental iterations. At the end of each second the pipe server writes an [8] x [3] array to the pipe containing the detector information over the last second. The array is then read by the pipe-client and the information is implemented in the modeled-world simulation. For a graphical representation of the experimental design, see Figure 4.

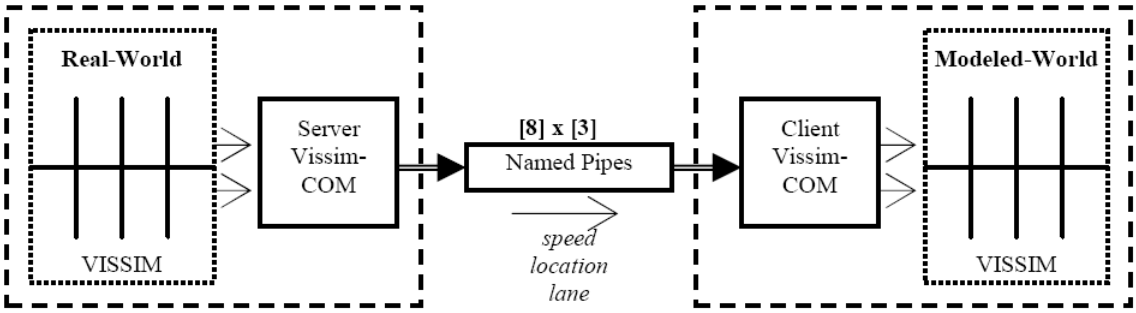


Figure 4 Experimental Design for Proof of Concept

The execution of the model world is driven by the real-world model, with the modeled-world executing a simulation second only when a second of data are received from the real-world data server. In this experiment, reliable, ordered communications are assumed with the named pipe operating on a first-in-first-out (FIFO) basis ensuring that the modeled-world and real-world simulations remain synchronized. Subsequent versions of the framework using the TRTI integrate timestamps directly into the data stream and

incorporate data consistence checks. The following pseudo-code further illustrates the structure of the server – client relationship.

```
Pipe-server (Real-World)
        for (i = 0, i <= simulation period, i++)
                {       advance simulation 1 sec
                        read  vehicle speeds from the 8 detectors
                        write [8]x[3] to pipe           }

Pipe-client (Modeled-World)
        for (i = 0, i <= simulation period, i++)
                {       read [8]x[3] from  pipe
                        input vehicle speeds into simulation
                        advance simulation 1 sec        }
```

4.1.1.1 Simulated Time Frame

A four hour simulation time period is used, capturing the transition into and out of the peak period. The flow rate is 500 vehicles per hour on the main arterial for the first hour, increasing steadily to 900 vehicles per hour over the second and third hours and then returning to 550 vehicles per hour in the fourth hour. At the end of the simulation period the average travel times and delays from seven representative paths, along with the queue lengths at each intersection approach are collected from both the real-world and modeled-world simulations (Figure 5 and

Table 4). These performance measures are presented in 10-minute interval aggregations.

4.1.1.2 Scenarios

The results from two scenarios are presented. Scenario 1 assumes ideal detector performance, with every real-world vehicle and its associated speed accurately detected and passed to the modeled-world. Under such an assumption the primary difference between the real-world and modeled-world results will be due to randomness in driver and vehicle characteristics and potentially different path selection decisions of a vehicle in the real-world and its simulated counterpart in the modeled-world. Scenario 2 introduces some of the variability expected in a field implementation from detector failures and speed measurement inaccuracies. The detectors randomly failed to detect vehicles with a frequency of approximately 2%. Additionally, the detected speeds were allowed to randomly vary higher or lower by up to 10% of the actual vehicle speed.

In both scenarios the vehicle speed measured over the detector in the real-world is used as the desired vehicle speed for the vehicle placed in the modeled-world. However, if the vehicle speed was lower than the expected range of desired speeds (48 to 58 kph) it is assumed the vehicle is within congested conditions and the desired speed is randomly set within the preceding desired speed range. In this instance the vehicle is placed in the modeled-world at the highest speed possible given traffic conditions without exceeding the desired speed. If the vehicle is traveling more slowly than its desired speed it will attempt to accelerate to its desired speed as quickly as possible.

4.1.2 Results and Analysis

Five replicate runs (R-1 through R-5) were performed. Each replicate run consisted of a modeled-world being driven by the streamed detector data of the real-world simulation, allowing for paired comparisons of the real-world and modeled world

simulations. Each replicate run utilized different random seeds for real-world and modeled-world simulation instances.

Travel time and delay results for seven paths and queue lengths for three approaches were compared between the real and modeled-world simulation instances for the two scenarios. Figure 5 presents the network link naming conventions and

Table 4 the performance measure links considered. All links in the network are single lane.

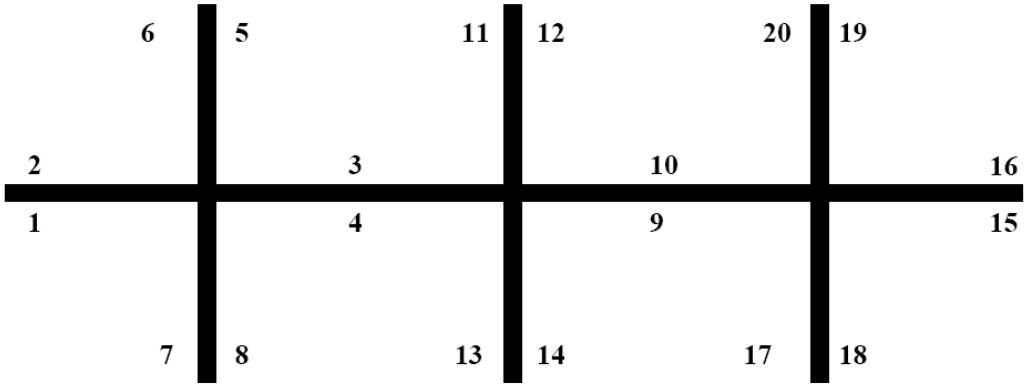


Figure 5 Roadway Network and Link Names

Table 4 Description of Performance Measures

Measures		Path Links	Distance (m)
Travel Time	Delay		
TT-1	DL-1	1-4-9-15	1308
TT-2	DL-2	16-10-3-2	1309
TT-5	DL-5	11-13	290
TT-8	DL-8	4-12	366
TT-9	DL-9	10-13	382
TT-10	DL-10	4-9	381
TT-11	DL-11	10-3	383
Queue Length			
QL-1		1	
QL-6		14	
QL-7		10	

4.1.2.1 Individual Performance Measures

For most of the monitored performance measures the Scenario 1 and Scenario 2 modeled-world simulations captured the performance of the real-world simulations accurately. For example, consider Figure 6 and Figure 7 which present the values of the travel time for path 2 (TT-2), from replication 2 (R-2), and of the queue length for path 6 (QL-6), from replication 3 (R-3), respectively, over the four-hour simulated time period. As seen, the modeled-world in both scenarios is able to reasonably track performance measures of the real-world through the four hour period.

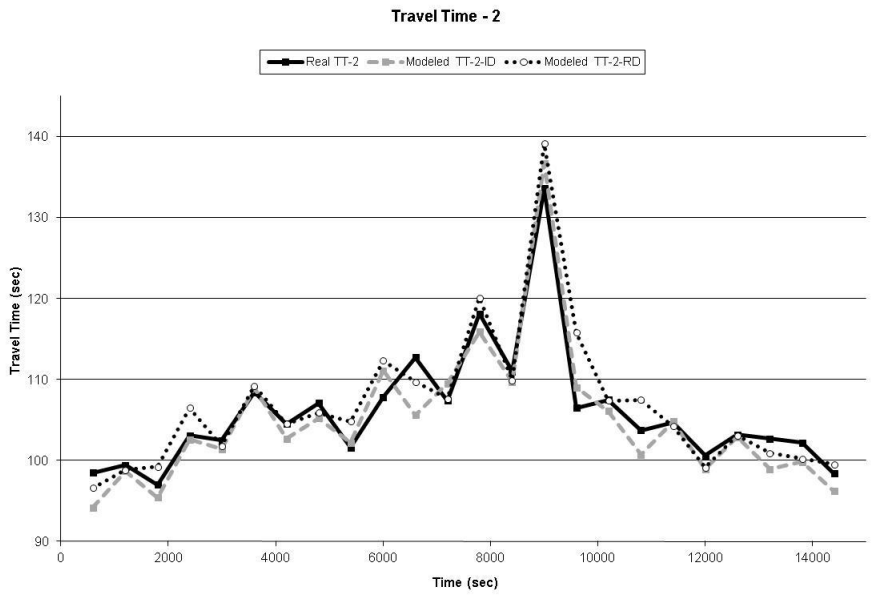


Figure 6 Average Travel Time for Travel Time Path 2 (TT-2), Replication 2 (R-2), for the Real-World and the Modeled World Scenarios 1 and 2.

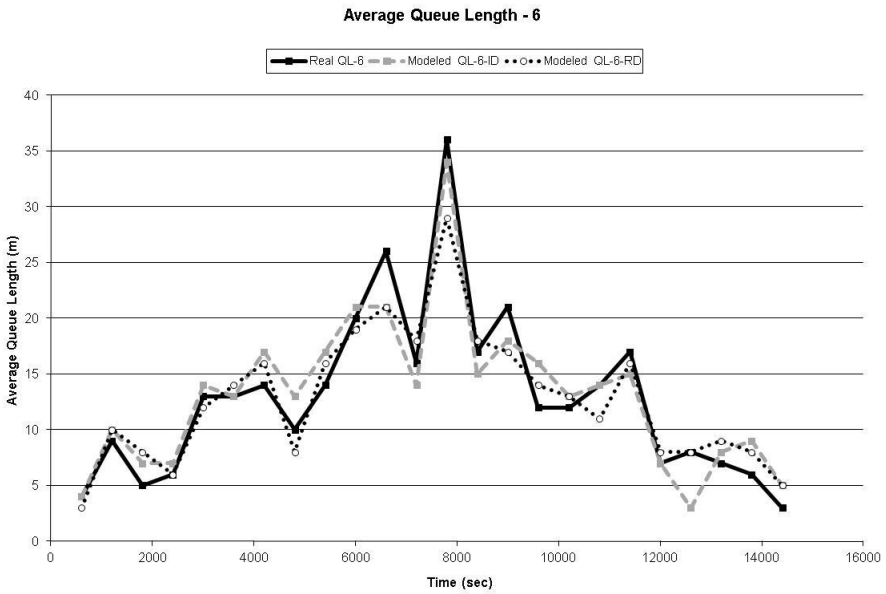


Figure 7 Average Queue Length for Queue 6 (QL-6), Replication 3 (R-3), for the Real-World, and the Modeled-World Scenarios 1 and 2

For instance, in Figure 6 and Figure 7 the absolute differences among the measures from the various scenarios analyzed are minimal. The average and standard deviation of the difference between the values of TT-2 from Scenario 1 are 2.06, and 1.55 seconds; and 2.10, and 2.12 seconds, respectively for Scenario 2. Similarly, the average and standard deviation of the difference between the values of QL-6 from Scenario 1 are 1.96, and 1.46 car-lengths; and 1.96, and 1.57, car lengths, respectively for Scenario 2.

However, when considering all replicated experiment instances it was found that the model-world did not always consistently track the real-world. For instance, consider Figure 8 and Figure 9, which represent the travel time for path 1 (TT-1) from replication 4 (R-4) and the delay for path 1 (DL-1) from replication 3 (R-3), respectively. There is a large discrepancy in the estimates of these particular performance measures between 8000 and 11000 seconds, the highest demand period of the simulated time frame. The modeled world travel time estimate approximately 73% of the real world travel time for

both Scenario 1 and Scenario 2. The delay estimate from the modeled world is approximately 44% of the estimate from the real-world.

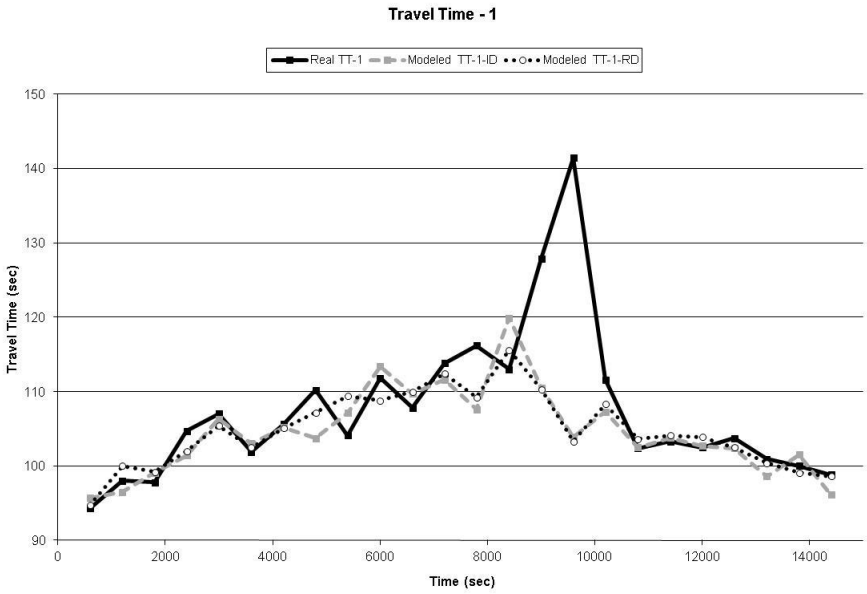


Figure 8 Average Travel Time for Travel Time Path 1 (TT-1), Replication 4 (R-4), for Real-World and Modeled-World Scenarios 1 and 2

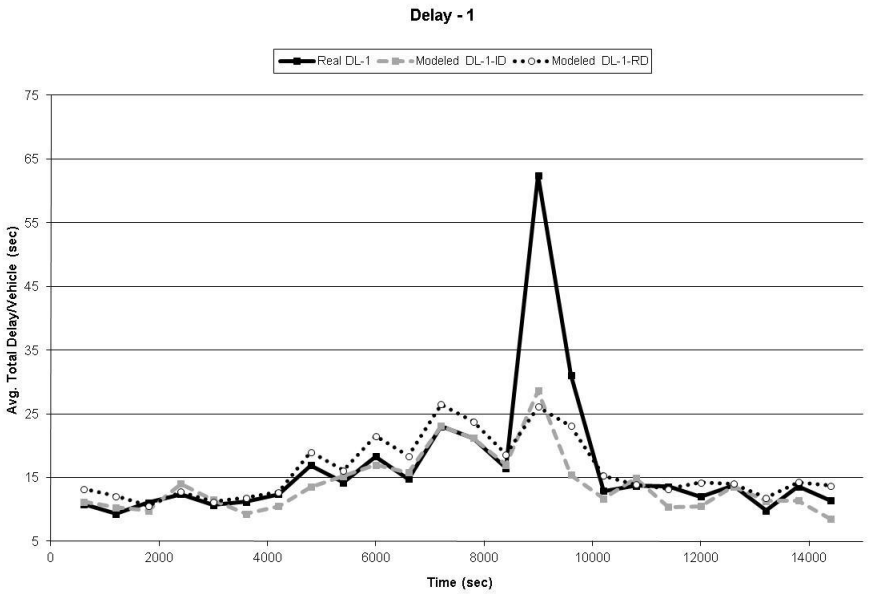


Figure 9 Average Delay for Approach 1 (DL-1), Replication 3 (R-3), for the Real-World, the Modeled-World Scenarios 1 and 2

Two potential sources of these errors are: 1) randomness in vehicle routing and 2) a smoothing of flows in the model-world. In these experiments the randomness in vehicle routing is limited to a vehicle's turning movement selection at an intersection. Of primary concern is the selection between through and left turn movements. For example, the intersection midway through the arterial has the highest left turn movement percentage at 16% in each direction. The impact of the randomness in left turn movement selection is seen through which vehicles in a particular platoon turn left. The left-turn vehicle placement in the queue can dramatically impact operations as flows approach capacity, particularly in this study network as a left-turning vehicle waiting for a gap will block all following (left, through, or right turning) vehicles. For example, if the 1st vehicle in a platoon is attempting to negotiate a left turn at the arterial's middle intersection and is unable to do so the waiting delay is incurred not only for the turning vehicle but also for

those vehicles queued behind the turning vehicle. Should the last vehicle in the platoon attempt to make a left turn, any delay while waiting for a gap will be experienced only by that left-turning vehicle.

This particular source of error cannot be addressed by boundary point sensors without knowledge of every real-world vehicle's desired path through the network. The current data-driven simulation is based on the hypothesis that such data are likely to be unavailable, at least in the near future. However, detector data from internal network detectors may provide a means to address this issue. For example, a mainline detector at the stop-bar could be used to identify when vehicles are not moving during a green phase and this information could be passed to the modeled-world simulation. The use of internal network detector information will be one direction of future research efforts.

The second issue, a smoothing of flows in the modeled-world, has the potential to "smooth" out traffic fluctuations. Currently, irrespective of the headway with which cars enter the real-world, the modeled-world implementation algorithm has the effect of rounding the headway to the nearest second. This is particularly noteworthy for actuated traffic control, where a few tenths of a second can be the difference between a signal gaping out and a car receiving an extension of the green phase. For example, in the replicate runs where the divergence in travel time was seen at the middle intersection it was also noted that the side streets tended to receive slightly more green time. Overall this would decrease the time given to the mainline and decrease the modeled-world delays. Headway smoothing of the entering flows is a likely explanation of the extended side street green time. Future efforts will consider methods to eliminate this unintended bias.

4.1.2.2 Consistency of Results

The consistency of the performance measures across replicate runs was explored by calculating the difference between the real-world and modeled world performance measures for each replicate trial and then averaging over the 5 replicates. Figure 10 and Figure 11 illustrate the concept of stability using average differences in queue lengths from Scenario 1 and travel times from Scenario 2, respectively.

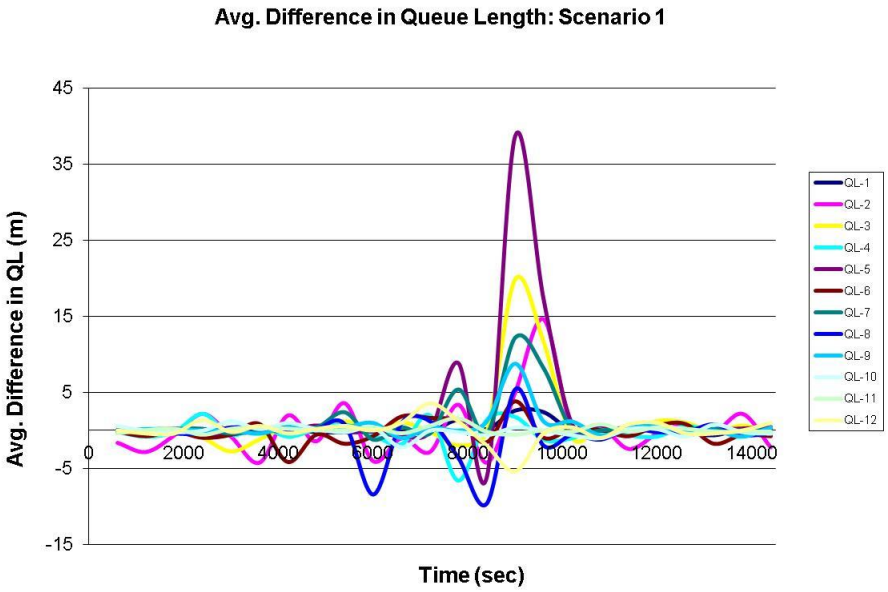


Figure 10 Average Difference in Queue Length, Scenario 1

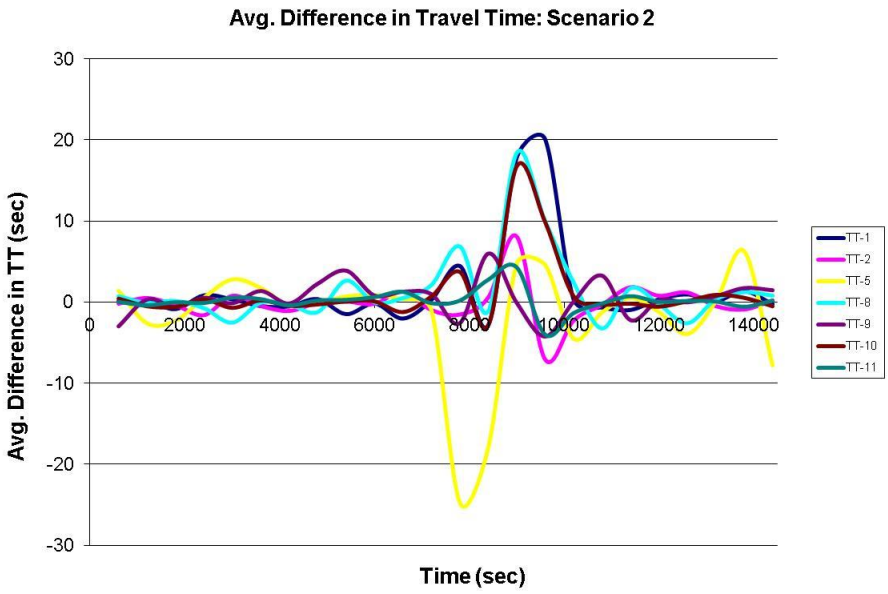


Figure 11 Average Difference in Travel Time, Scenario 2

The above figures indicate that the methodology being considered is rather stable except for a few instances were performance measures differed during the period being considered in a particular replicate trial. The reason for these differences is discussed in the previous section and will be the focus of future efforts. Overall, the modeled-world is generally successful at replicating performance measures of the real-world. In addition the method is seen to be resilient to reasonable detection errors, that is, drastically faulty data or complete detector failure is not considered in this analysis.

4.1.3 Limitation and Future Direction

In designing the proof of concept experiment, limited data were passed from the real-world simulation to the modeled-world simulation to data that could be obtained in a field implementation. That is, the modeled-world was not provided with more information than may be detected on today's roadways. However, in VISSIM there are

approximately 12 potentially influential parameters that are used for the purposes of calibrating traffic simulation models. Table 5 lists these parameters [40]. In the discussed experiment these 12 parameters are the same in the real-world and modeled-world simulations. This results in the modeled-world simulation having "perfectly" calibrated parameters relative to the real-world.

Table 5 Description of VISSIM Calibration Parameters [40]

Parameters	
Emergency stopping distance	Minimum headway
Lane changing distance	Desired safety distance parameters
No. observed preceding vehicles	Maximum deceleration
Maximum look ahead distance	-1 m/s^2 per distance
Average stand still distance	Accepted deceleration
Waiting time before diffusion	Distance of standing and 50km/h

One of the key next steps is the exploration of the impact of these calibration parameters and other sources of randomness in the simulation. Chapter 5 presents an in depth discussion on model calibration.

Finally, the current model is limited to detection at boundary points of the model. Future work will seek the incorporation of detection data from internal detectors into the model calibration. This will consider standard detections (i.e. typical actuated control layouts) and the possible of new detector placement specifically designed to aid a real-time simulation.

4.1.4 Experiment #1 Summary

This experiment explored a methodology to develop a data-driven online simulation tool to deliver real-time performance measures with the aid of microscopic

traffic simulation. The major objective of this experiment was to demonstrate the feasibility of such as real-time simulation. A proof of concept experiment was designed to have one microscopic traffic simulation instance reflect the performance measures of second model, using only data that could be polled from a detector. This experiment was accomplished through the use of two VISSIM simulation instances, where one represented the real-world and the other, the modeled-world.

The results from this experiment demonstrated that the modeled-world is capable of reflecting the performances measures of the real-world with a relatively high level of accuracy. However, some notable discrepancies were seen. Despite the current discrepancies and limitations of the experimental design, the results presented support the likely feasibility of this approach.

4.2 Experiment #2: Field Test with Temporary Detectors

In experiment #1, the results of preliminary studies to determine the feasibility of the proposed framework are presented. Given the feasibility of the proposed methodology in a simulated environment, a field test was developed to explore the methodology's robustness. The goal of the initial field test was, in part, to determine whether a VISSIM simulation instance could be driven by real-time, real-world, detector data and produce performance measures that reflect those of the area being simulated. To conduct this experiment, the 5th Street / Ferst Drive corridor in the midtown Atlanta area on the Georgia Tech campus was selected as the arterial to be studied (see Figure 12). The experiment was conducted for 90-minutes, during the peak noontime period on July 16, 2009.

A VISSIM model for the test bed area was developed. Real-time detector data was streamed into the VISSIM simulation model from boundary detectors (Figure 13). While not used in this experiment midblock detector data was also streamed and logged for potential use in future concept development efforts. For this experiment temporary detectors where utilized. Detector data was transmitted over Georgia Tech's wireless network to a central data processing server. A time stamped message was sent for each vehicle that crossed a detector. The time stamped data included the link number and lane number of the reporting detector and the measured vehicle speed. In addition, the corridor was outfitted with temporary cameras located at each of the six intersections that record arterial operations during the experiment. The cameras facilitate the post-hoc extraction of travel time data to be used in the evaluation of the real-time simulation performance. In addition, two GPS equipped vehicles logged their location and speed data as they traversed the study corridor during the experiment. Figure 13 shows the VISSIM representation of the test site and the locations of detectors and cameras along the 5th Street NW and Ferst Drive NW corridor. At the end of the 90-minute test period the logged data was processed and various performance measures extracted for comparison with the simulation output.

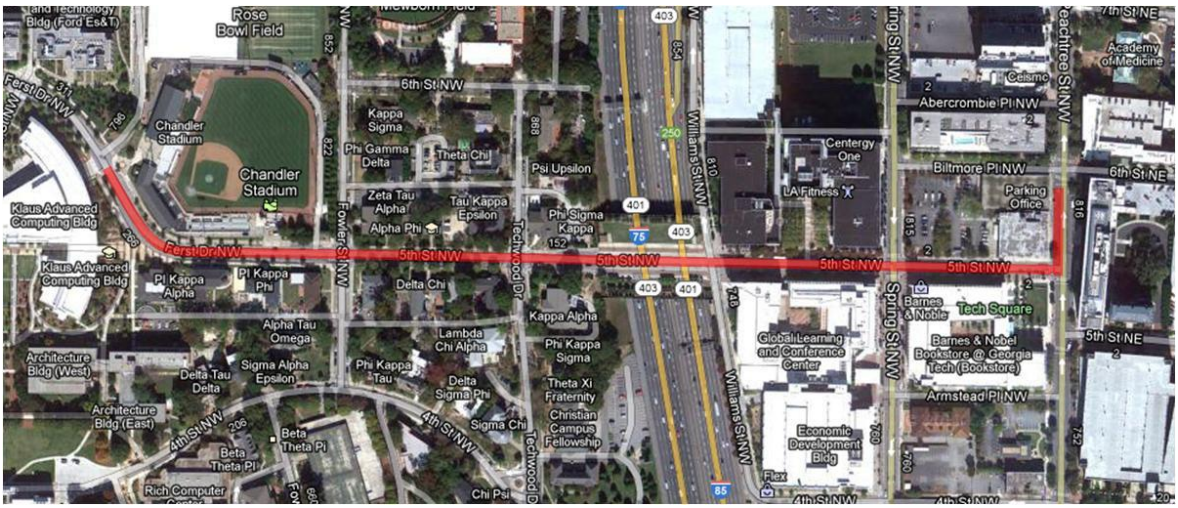


Figure 12 5th Street NW/Ferst Drive NW Study Corridor (red line), Atlanta GA

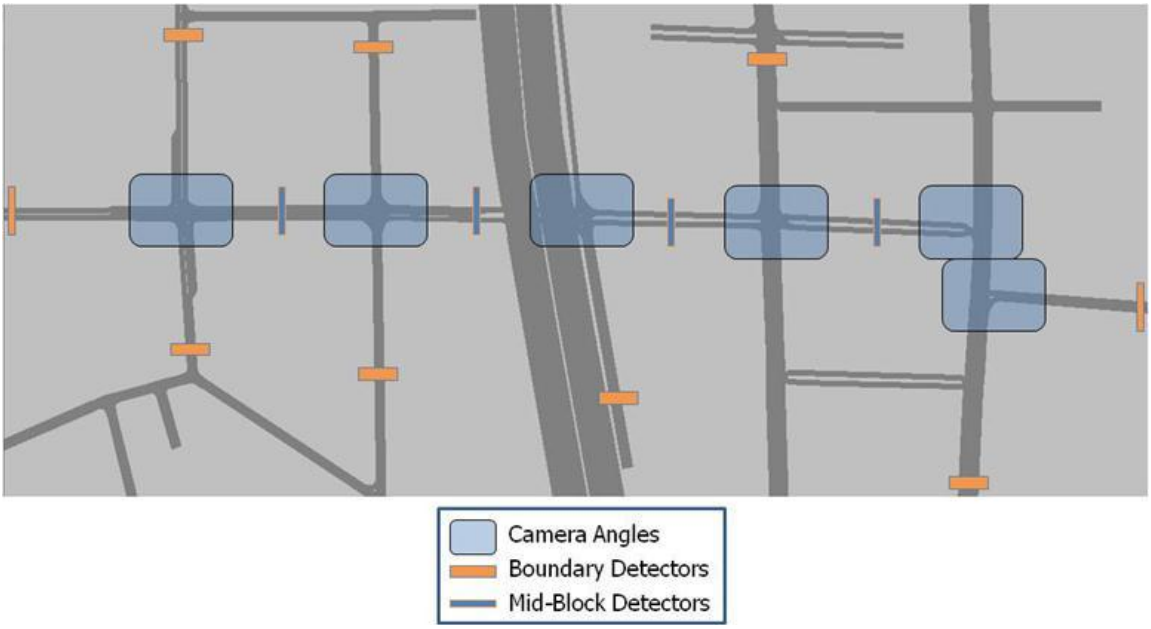


Figure 13 VISSIM Representation of the Study Corridor

4.2.1 Results and Analysis

The camera and video based field travel time data were compared with the VISSIM model to determine how well the data-driven simulation was able to reflect field

travel times. Two primary sets of travel times were obtained: eastbound (EB) and westbound (WB). Scatter plots of the data are shown in Figure 14 and

Figure 15. For the eastbound data, Figure 15, one can readily infer that the VISSIM travel times are similar to the field travel times, with exceptions at the boundaries of the graphic where the VISSIM travel times appear to be higher than the field travel times. For the westbound data sets, Figure 14, there is less similarity between the VISSIM and field travel times. From the westbound graphic the field travel times appear systematically in the lower range of travel times output by VISSIM.

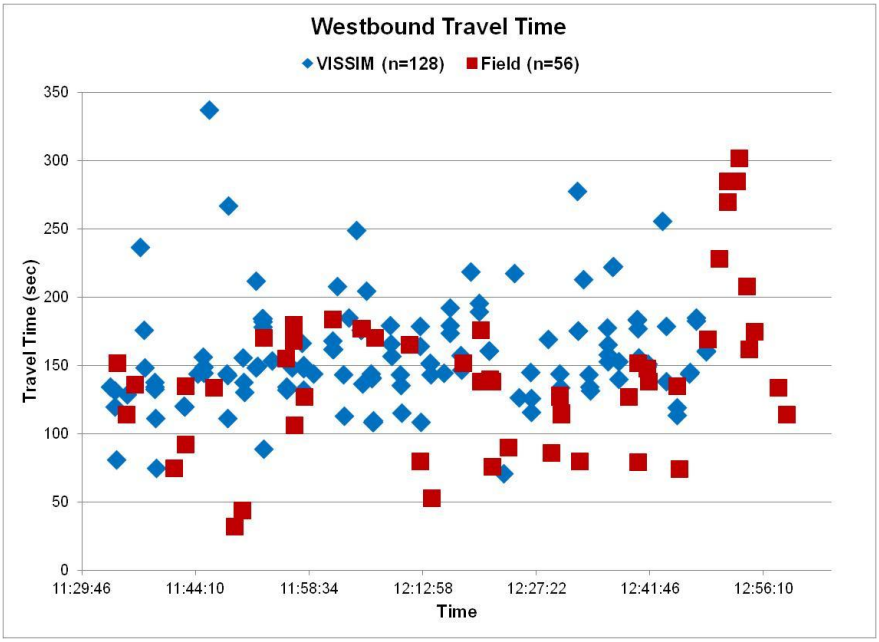


Figure 14 Westbound Travel Times - VISSIM vs. Field

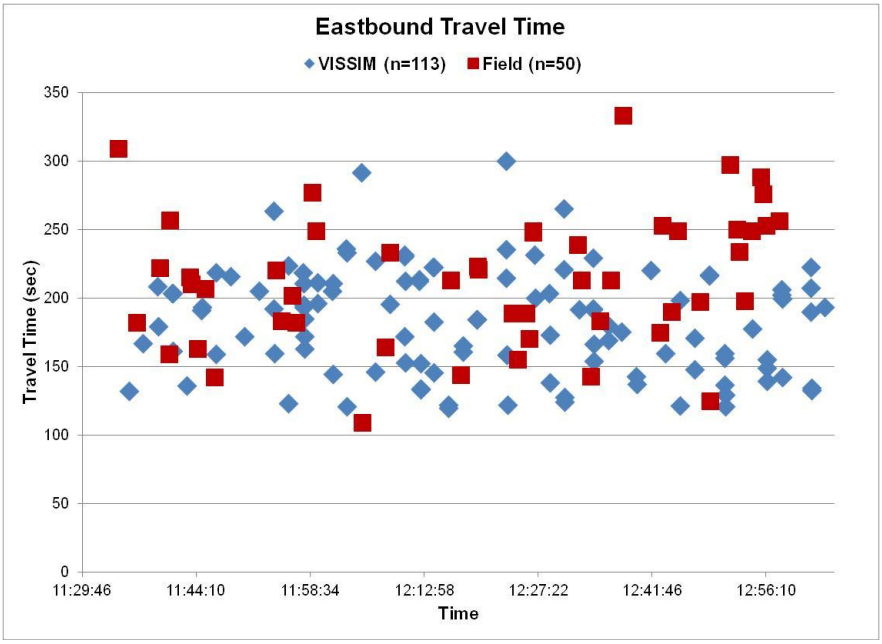


Figure 15 Eastbound Travel Times VISSIM vs. Field

The descriptive statistics were next examined (Table 6). In the eastbound dataset, the VISSIM travel times have a mean of 183.1 seconds and a standard deviation of 39.3 seconds. The field measured travel times have a mean of 218.4 seconds and a standard deviation 50.1 seconds. The higher eastbound field measured travel time does not appear to be systematic but heavily influenced by a cluster of high values near the end of the run. An analysis of the data removing the last fifteen minutes reduces the difference in average travel time between the Eastbound simulated and field results by approximately 45% percent, from a travel time of 218.4 to 202.3 seconds. Potential reasons for this cluster will be discussed later in the section. For the westbound direction, the mean and standard deviation of the VISSIM travel times are 157.5 seconds and 38.9 seconds respectively, while the field measured travel times the mean and standard deviation are 113.4 seconds and 63.0 seconds, respectively.

Table 6 Descriptive Statistics for Eastbound and Westbound Travel Times

Statistic	Eastbound Travel Time		Westbound Travel Time	
	VISSIM	Field	VISSIM	Field
Mean	183.1	218.4	157.5	113.4
Standard Deviation	39.3	50.1	38.9	63.0

Next, statistical tests were conducted to determine whether the VISSIM and the field measured travel times are statistically different. First the distributions were tested for normality as this will influence the statistical test chosen. Lilliefors normality tests were conducted on all the travel time data sets. The results of the normality tests are presented in Table 7. From these results one is unable the reject the null hypothesis that the eastbound VISSIM and Field travel times are normally distributed. However for westbound VISSIM and Field travel times the normality test results provides sufficient evidence for one to reject the assumption that these datasets are normally distributed. These conclusions were further corroborated after examining a series of density plots, Figure 16 and Figure 17, and Q-Q plots, Figure 18 and Figure 19.

Table 7 Statistical Test Results

Statistical Test	p-Value	Interpretations
Normality Test		
EB VISSIM	0.3255	Unable to reject normality assumption
WB VISSIM	0.0001	Reject normality assumption
EB Field	0.6760	Unable to reject normality assumption
WB Field	0.0088	Reject normality assumption
2 Sample t-Test		
EB VISSIM vs. EB Field	0.0001	Reject equal mean assumption
WB VISSIM vs. WB Field	0.1125	Unable to equal mean assumption
Wilcoxon Sum Rank Test		
EB VISSIM vs. EB Field	0.0001	Reject equal median assumption
WB VISSIM vs. WB Field	0.0408	Reject equal median assumption
Chi-Square Test		
EB VISSIM & EB Field	0.3654	Unable to reject same distribution assumption
WB VISSIM & WB Field	0.1560	Unable to reject same distribution assumption
Kolmogorov-Smirnov Test		
EB VISSIM & EB Field	0.0016	Reject same distribution assumption
WB VISSIM & WB Field	0.0235	Reject same distribution assumption

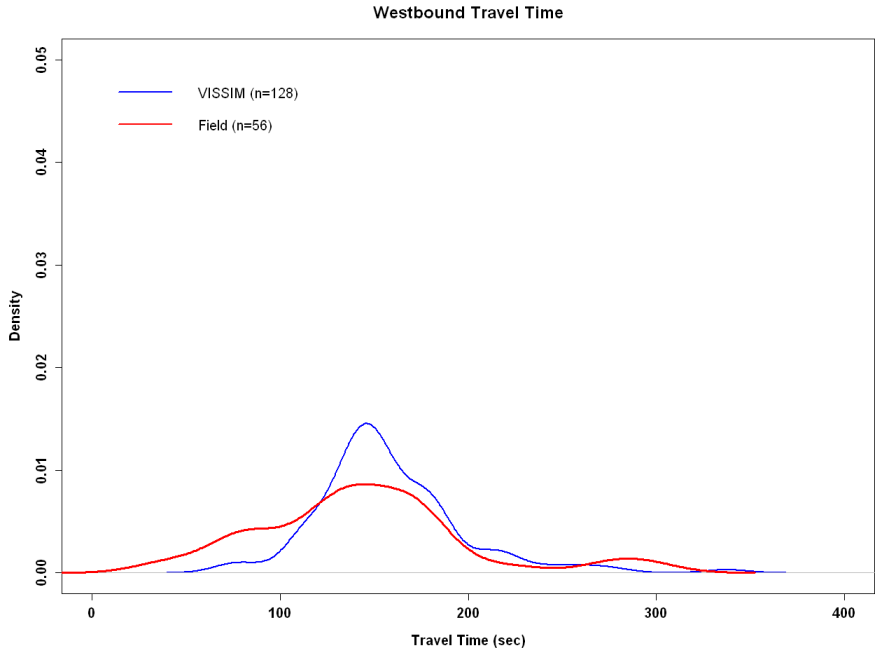


Figure 16 Density Plot of VISSIM vs. Field Westbound Travel Times

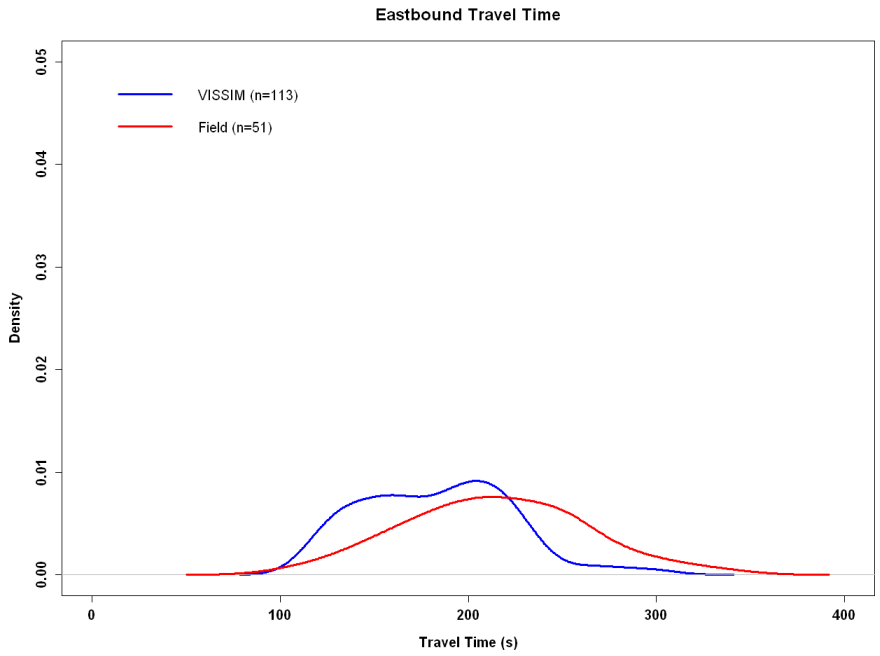


Figure 17 Density Plot of VISSIM vs. Field Eastbound Travel Times

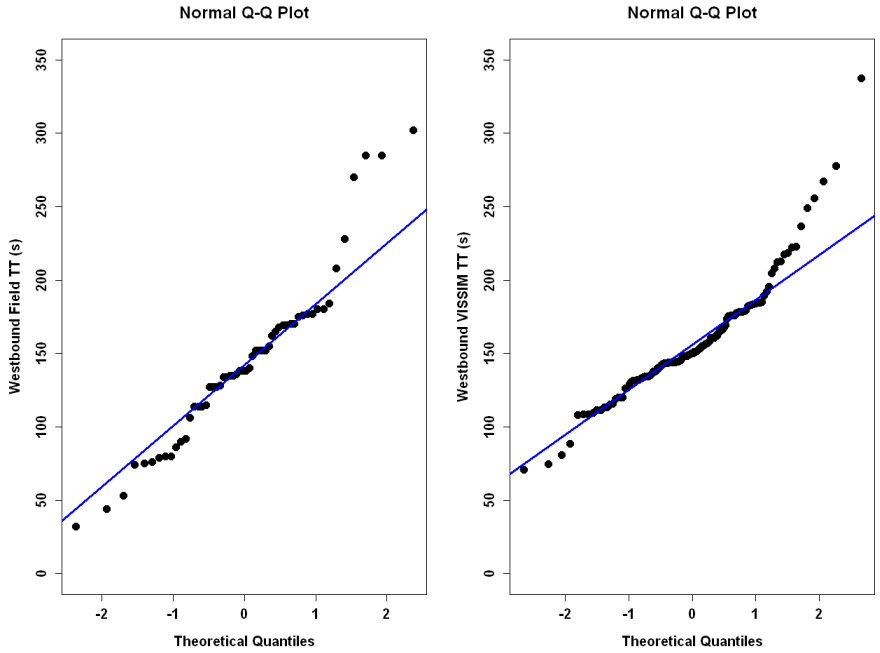


Figure 18 Q-Q Plots of Field and VISSIM Westbound Travel Times

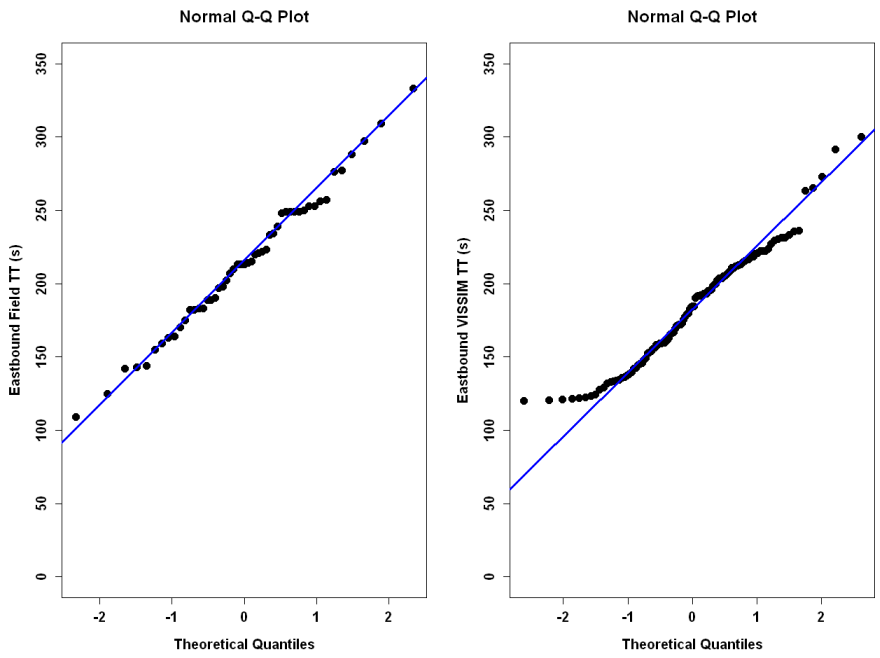


Figure 19 Q-Q Plots of Field and VISSIM Eastbound Travel Times

A series of other statistical tests were conducted to further explore the differences between Field and VISSIM travel time estimates. These tests were also used to quantify some of the similarities and dissimilarities that were observed, especially from the density plots. The test results are also included in Table 7.

From the above results one can conclude that there is a statistical difference between the VISSIM and the actual (mean / median) travel times, in both the eastbound and the westbound directions. However, it is again noted that if the last fifteen minutes of data were not included in the eastbound analysis the result is reversed, with the test failing to reject equal means. This further indicates an event specific issue rather that a systemic problem eastbound.

Several areas are perceived as potential sources for the discrepancies between the estimated and actual performance measures. These areas are generally related to model

calibration and facility representation. For model calibration the parameters that reflect driver behavior were left unchanged in VISSIM, potentially indicating that the default driver behavior in the VISSIM model may not be representative of the behavior along the study corridor. Accurately capturing driver behavior may improve VISSIM estimates of travel times. In addition, considerable differences in simulated vs. field volumes were observed on some links. In part this is a result of simulated turning movement distributions at the various intersections throughout the corridor differing significantly from the field movements. Historical turning movement percentages where utilized as real-time turning counts were not available. It is also noted that volume discrepancies could result from detector errors. The Tech Trolley (an on-campus shuttle) was also not represented. By not capturing the Trolley behavior, VISSIM is not able to simulate the increase in travel time for other vehicles that the Trolley may inhibit as it traverses the corridor.

There are two aspects of the study corridor that were not represented in the VISSIM model of the area. The first was the roadway gradient, which is positive from west to east, and the second, the pedestrian and pedestrian facilities along the corridor. Thus, any influence from these factors is not reflected in the VISSIM model. Pedestrians in particular were noted as a potential significant factor. The probe vehicle drivers noted instances where pedestrian movements significantly interfered with traffic flow. For example, at the intersection of 5th Street and Spring Street left turning vehicles yielding to crossing pedestrians would prevent through vehicles behind the left turning vehicle from traversing the intersection. As no pedestrians were modeled in VISSIM this behavior was not reflected. For the test bed the pedestrian interference was particularly notable as given

the nature of a college campus there tends to be periods in which significant, short duration, increases in pedestrian activity occur.

4.2.2 Experiment #2 Summary

In experiment #2, the fundamentals of the proposed methodology were explored in a field test using readily available technology. The microscopic traffic simulation model was able to be driven in real-time by real-world data streams. The comparative analysis demonstrated some success particularly when considering the eastbound travel times. It is anticipated that once sources of identified discrepancies are addressed the VISSIM model will be able to produce better estimates of travel times.

4.3 Experiment #3: Field Test with Temporary and Permanent Detectors

Test #3 is a full scale test of the methodology. The field test was conducted on July 8, 2010, between 1:00PM and 3:00PM. The study area is the same as the previous test (Figure 12). Both permanent (Video Detection System (VDS)) and temporary detectors, capable of streaming individual vehicle records, were employed during this test. In addition to the temporary and permanent detectors six camcorders were used to collect additional traffic information for post processing. Four camcorders were used to detect boundary conditions (i.e. when vehicles enter and exit the network) while two were used to collect signal phase information at the intersections of 5th Street and Spring Street, and 5th Street and W. Peachtree Street. The location of each detector, and their respective detection zones, including the camcorders and their view angles are shown in Figure 20. In addition, probe vehicle travel routes were added to allow for a more robust

evaluation of the system. In this field test four routes are monitored, each of which are traversed by two probe vehicles, see Figure 21.

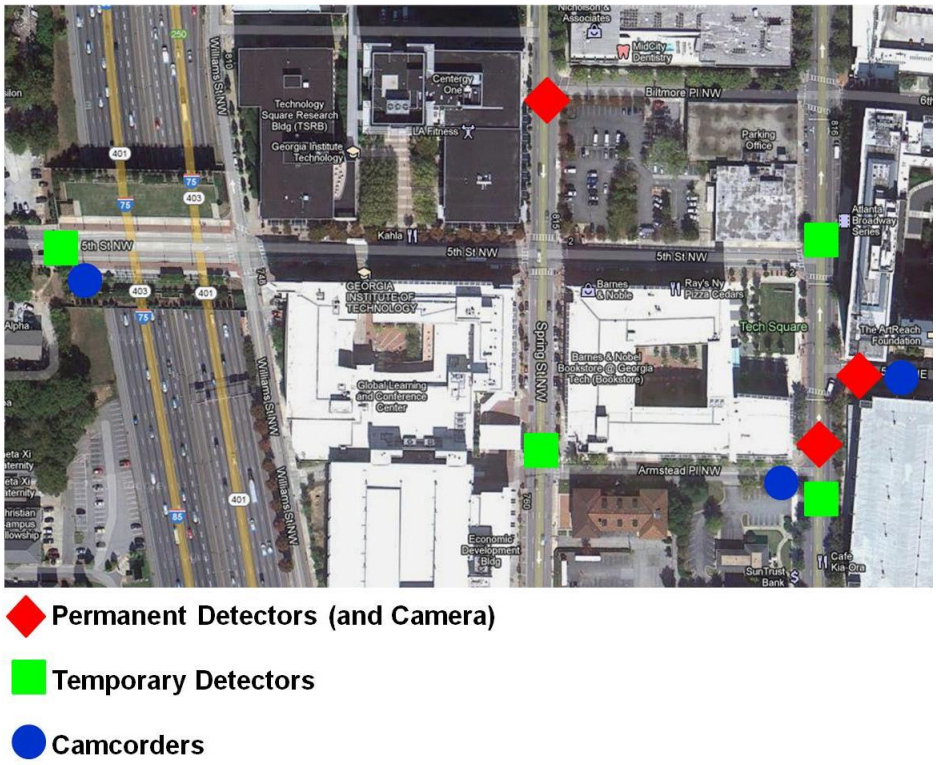


Figure 20 Detector Locations Throughout the Study Area

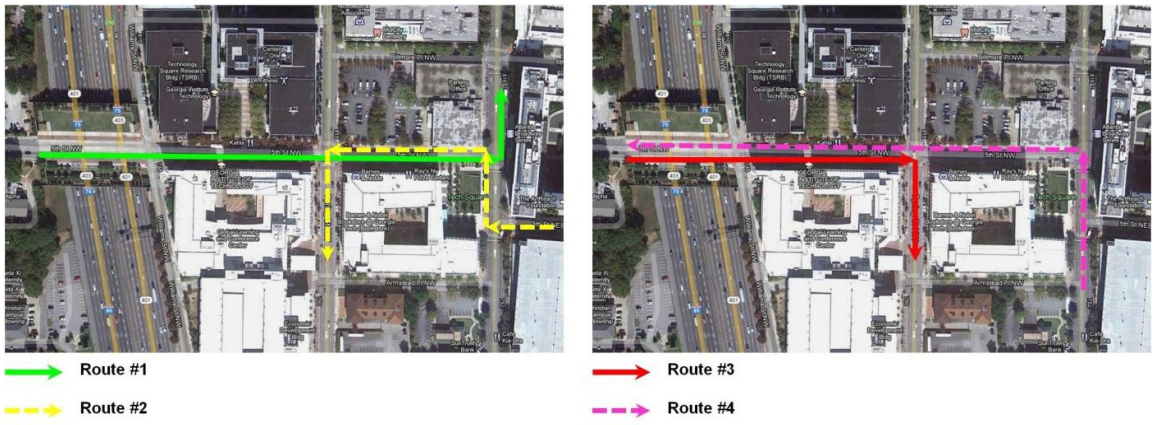


Figure 21 Probe Vehicle Routes Through Study Area

Each permanent VDS camera is connected to the Georgia Tech fiber network and is capable of detecting and transmitting individual vehicle records in real-time. The temporary detectors consisted of research assistants with laptop computers using a script to record and transmit individual vehicle data back to the server in the laboratory. These detectors were primarily tasked with detecting the four probe vehicles. By identifying the probe vehicles in the field in real-time they could be identified as they entered the simulation, allowing for a more robust paired travel time comparison in the later analysis. The temporary detector on the 5th Street bridge was also tasked with detecting non-probe vehicles as a permanent VDS camera was not available for this site.

Each packet of transmitted detector data includes six fields. They are detector number (each detector location having a unique number), lane number, speed (in miles per hour), detector time, and epoch time. Table 8 provides a sample of streamed data. Clock time is also presented in the sample below but it is determined from the epoch and not transmitted by the detectors.

Table 8 Sample of Streamed Detector Data

Detector #	Lane #	Speed (mph)	Timestamp	Epoch Time	Clock Time
4	2	18	13:00:45	1278608487.375490	13:01:27
11	1	22	11:04:17	1278608487.578350	13:01:28
11	3	8	11:04:17	1278608487.677290	13:01:28
11	2	6	11:04:17	1278608487.779180	13:01:28
10	1	17	13:01:25	1278608487.935580	13:01:29
5	1	26	13:00:20	1278608487.200850	13:01:29
1	1	6	12:57:53	1278608487.419210	13:01:29
11	3	9	11:04:19	1278608487.778030	13:01:30

During a preliminary test, videos feeds were compared to VISSIM animation to verify that as a vehicle entered a detection zone the detector data was successfully

transmitted and VISSIM generated a vehicle in the appropriate position. Figure 22 is an example image from the verification process.

Figure 22 Permanent Detector Generating Vehicle in VISSIM in Real-Time

Data was collected for approximately 120 minutes. At the end of this data collection, six different data streams were available:

- GPS data from the 4 probe vehicles
- Signal phase information from the two signalized intersections
- Vehicle presence from permanent video detectors
- Probe vehicle presence from temporary detectors
- Individual vehicle travel times over the pre-defined routes
- VISSIM trajectory data for all vehicles generated from the arriving data stream

4.3.1 Results and Analysis

Where the previous tests focused on aggregate travel times this experiment sought to evaluate the real-time simulations ability to estimate an individual vehicle's travel time through paired travel time comparisons. Probe vehicle travel times were extracted from the video footage and from the simulation's equivalent vehicle. These two sets of travel times were then compared.

The following discussion focuses on travel times for probe vehicles traveling along routes #2 and #4 as similar inferences can be made from the analysis of the data from routes #1 and #3. Route #2 is approximately 1300 feet in length and traverses three signalized intersections (Figure 21). Route #4 is approximately 1600 feet in length and includes three signalized intersections.

Twenty four pairings of travel times were collected from Route #2 and 36 from Route #4. Each pairing consists of a field probe vehicle travel time and the respective simulation estimate. The average field travel time for Route #2 is approximately 94 seconds and the simulation estimate is approximately 85 seconds. The Route #4 field travel time estimate is approximately 136 seconds and the simulation estimate is approximately 121 seconds. These and other descriptive statistics can be seen in Table 9. Figure 23 and Figure 24 present scatter plots of individual travel time estimates. Figure 25 and Figure 26 are also included to present travel time data from Route #1 and Route #3. Figure 27 presents four pairs of density plots to further compare each pair of travel time estimates for each route.

Table 9 Descriptive Statistics of Travel Times for Route #2 and #4

	Mean Travel Time		Standard Deviation Travel Time	
Route #	VISSIM	Field	VISSIM	Field
1	138.6	141.5	20.9	20.6
2	84.8	93.5	27.7	27.1
3	42.8	45.65	23.0	31.8
4	121.3	135.6	30.6	27.2

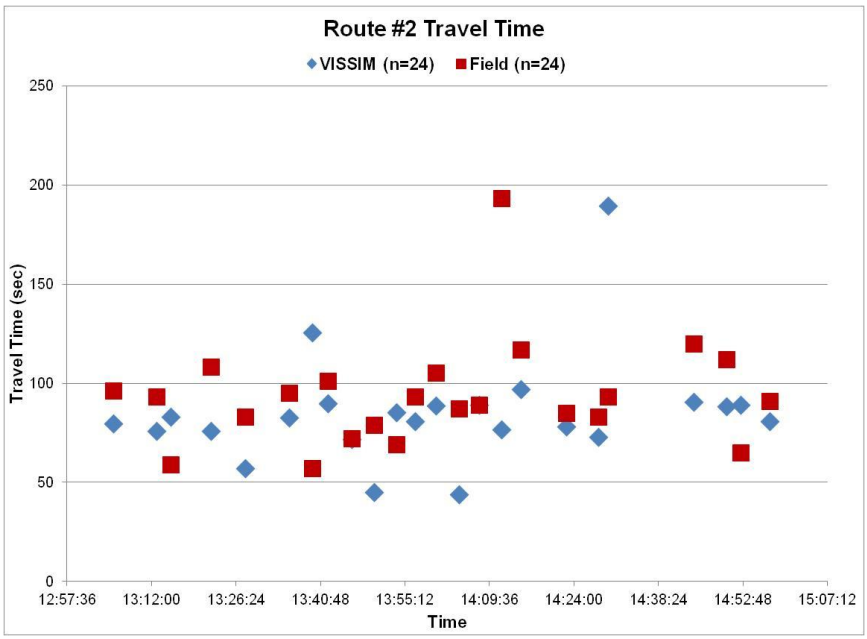


Figure 23 Route #2 Travel Times – VISSIM vs. Field

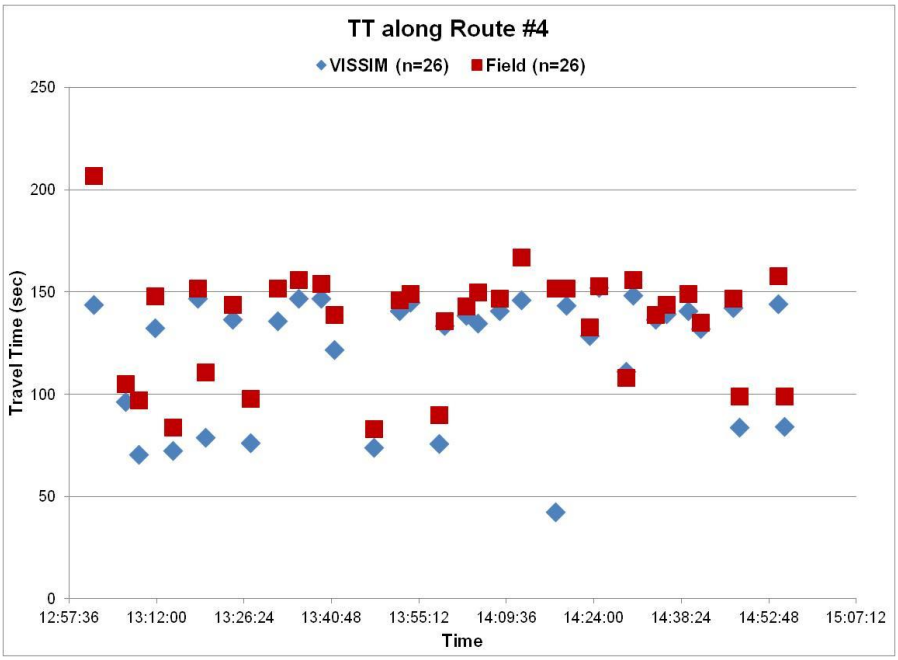


Figure 24 Route #4 Travel Time – VISSIM vs. Field

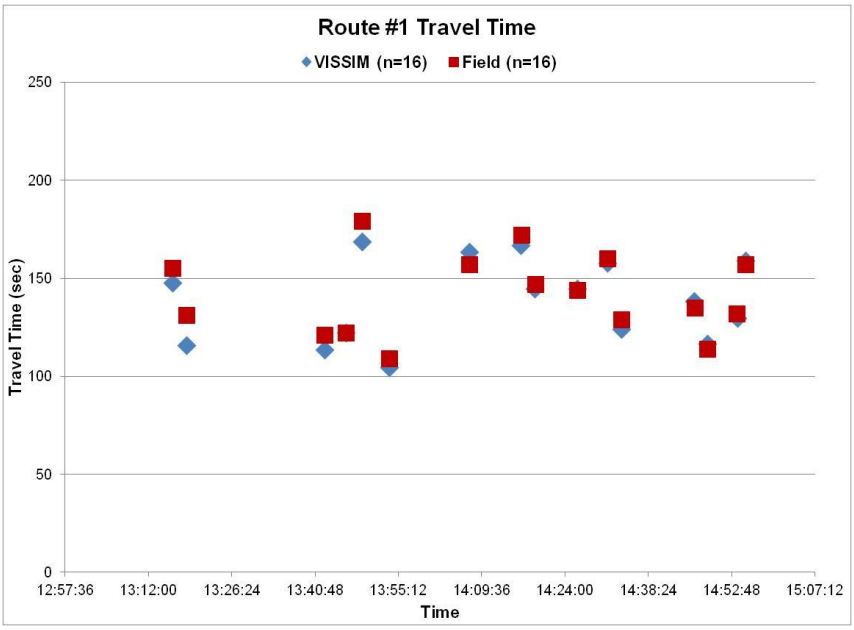


Figure 25 Route #1 Travel Time – VISSIM vs. Field

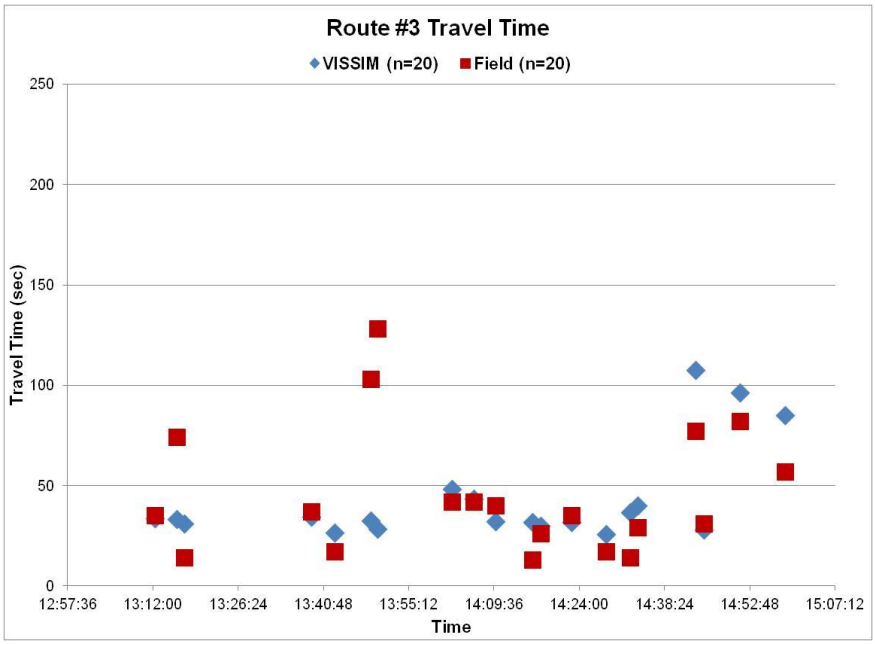


Figure 26 Route #3 Travel Time – VISSIM vs. Field

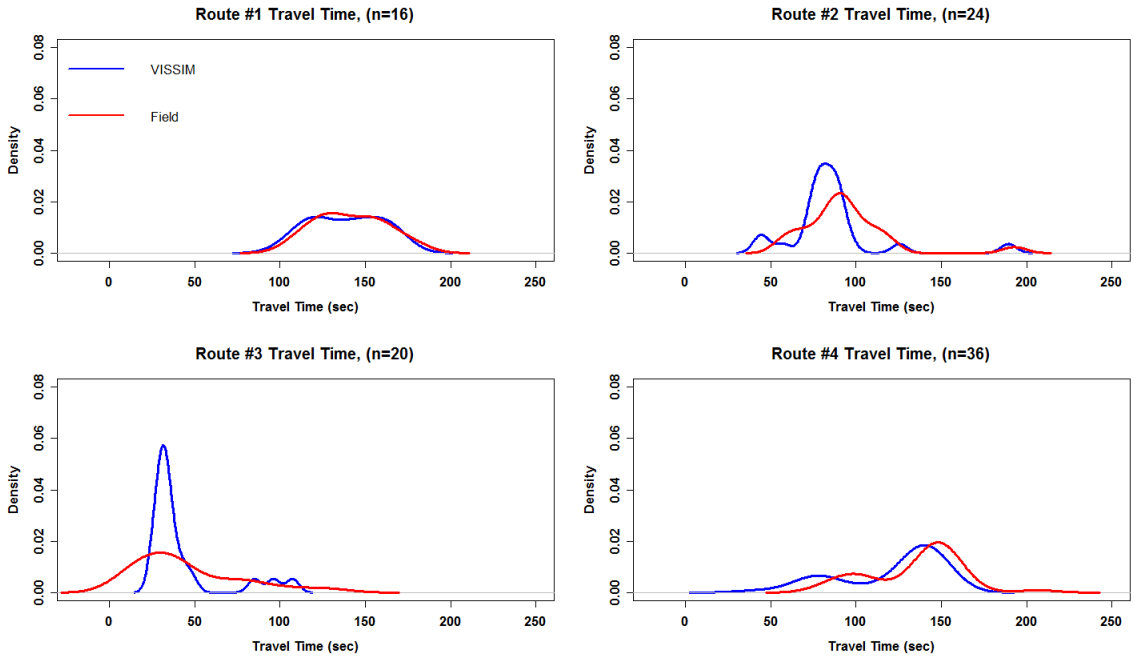


Figure 27 Density Plots of VISSIM and Field Travel Times for Routes #1 - #4

Similar to the previous analysis a series of statistical tests was conducted. The conducted tests include a paired t-test and a sum rank test, to compare means/medians, and chi-square test and a Kolmogorov-Smirnov test, to compare distributions. The results of these tests are presented in Table 10.

Table 10 Test Statistics

Statistical Test	p-Value	Interpretations
Normality Test		
Rt. #2 VISSIM	0.00001	Reject normality assumption
Rt. #4 VISSIM	0.00000	Reject normality assumption
Rt. #2 Real-World (RW)	0.06387	Unable to reject normality assumption
Rt. #4 Real-World (RW)	0.00198	Reject normality assumption
2 Sample t-Test (Paired)		
Rt. #2 VISSIM vs. Rt. #2 RW	0.28670	Unable to reject equal mean assumption
Rt. #4 VISSIM vs. Rt. #4 RW	0.00013	Reject equal mean assumption
Wilcoxon Sum Rank Test		
Rt. #2 VISSIM vs. Rt. #2 RW	0.05382	Unable to reject equal median assumption
Rt. #4 VISSIM vs. Rt. #4 RW	0.00549	Reject equal median assumption
Chi-Square Test		
Rt. #2 VISSIM & Rt. #2 RW	0.28930	Unable to reject same distribution assumption
Rt. #4 VISSIM & Rt. #4 RW	0.15740	Unable to reject same distribution assumption
Kolmogorov-Smirnov Test		
Rt. #2 VISSIM & Rt. #2 RW	0.03101	Reject same distribution assumption
Rt. #4 VISSIM & Rt. #4 RW	0.00864	Reject same distribution assumption

Based on the scatter plot data and the statistical tests it may be concluded that the simulation reasonably reflects the real world however differences do exist. It is noted immediately that a significant improvement from the previous test was the synchronization the signal in the simulation with the field, likely accounting for much of the improved performance.

However, several issues may be readily noted when reviewing the analysis. Firstly, the simulated estimates of the probe vehicles' travel time tend to be lower than

the field measured travel time. One potential reason for this result is vehicle acceleration rate in the field versus the simulation. During the test run, vehicles in the simulation appeared to accelerate to their desired speeds more aggressively than vehicles in the field. Acceleration rates can be a significant factor, particularly in a network dominated by signalized intersections. These rates can determine whether a vehicle arrives on red or green at a downstream intersection, which directly affects travel time estimates as well as other performance measures. For instance, on several occasions it was observed that a simulated vehicle successfully traversed a downstream signal with the corresponding vehicle in the field arriving a few seconds later stopped at a red light. While differences in acceleration rates do not often lead to such dramatic differences, they also can lead to more subtle changes. This again highlights the need for underlying calibration of the simulation model.

There are a several other subtleties that may be contributing to the discrepancies in travel time estimates. As mentioned previously, three of the more significant contributing factors are signal synchronization, vehicular volume traversing the network, and turning movement distributions. In the preceding experiment, a methodology was developed to synchronize the signals in the simulated environment and the field. However, real-time methodologies are not yet available to address the other two issues. The next reported test attempts to remove these issues and explore the capabilities of a real-time simulation given (near) perfect data.

4.4 Experiment #4: NGSIM's Peachtree Corridor Study

4.4.1 Motivation and Background

Experiment #4 may be described as a pseudo field test. The experiment utilizes a near ideal data set (tenth of a second resolution of vehicle positions on the corridor, route data for every vehicle, individual vehicle turning movement data, and signal status at a tenth of a second resolution) to determine the performance of the real-time simulation under ideal conditions. That is, under ideal data collection conditions is a real-time simulation capable of providing a reasonable reflection of the real world. This experiment uses previously collected field data as input to the real-time simulation, streamed in wall clock time. This data was collected as part of the FHWA Next Generation Simulation (NGSIM) program [43]. The NGSIM program created high fidelity data sets intended for use in the study of traffic behavior and the development of the next generation of traffic simulation tools and algorithms.

The NGSIM data set utilized is for Peachtree Street, Atlanta GA. The studied segment extended just south of the intersection of Peachtree and 10^{th} Street and north of the intersection of Peachtree and 14^{th} Street, Figure 28. [43]

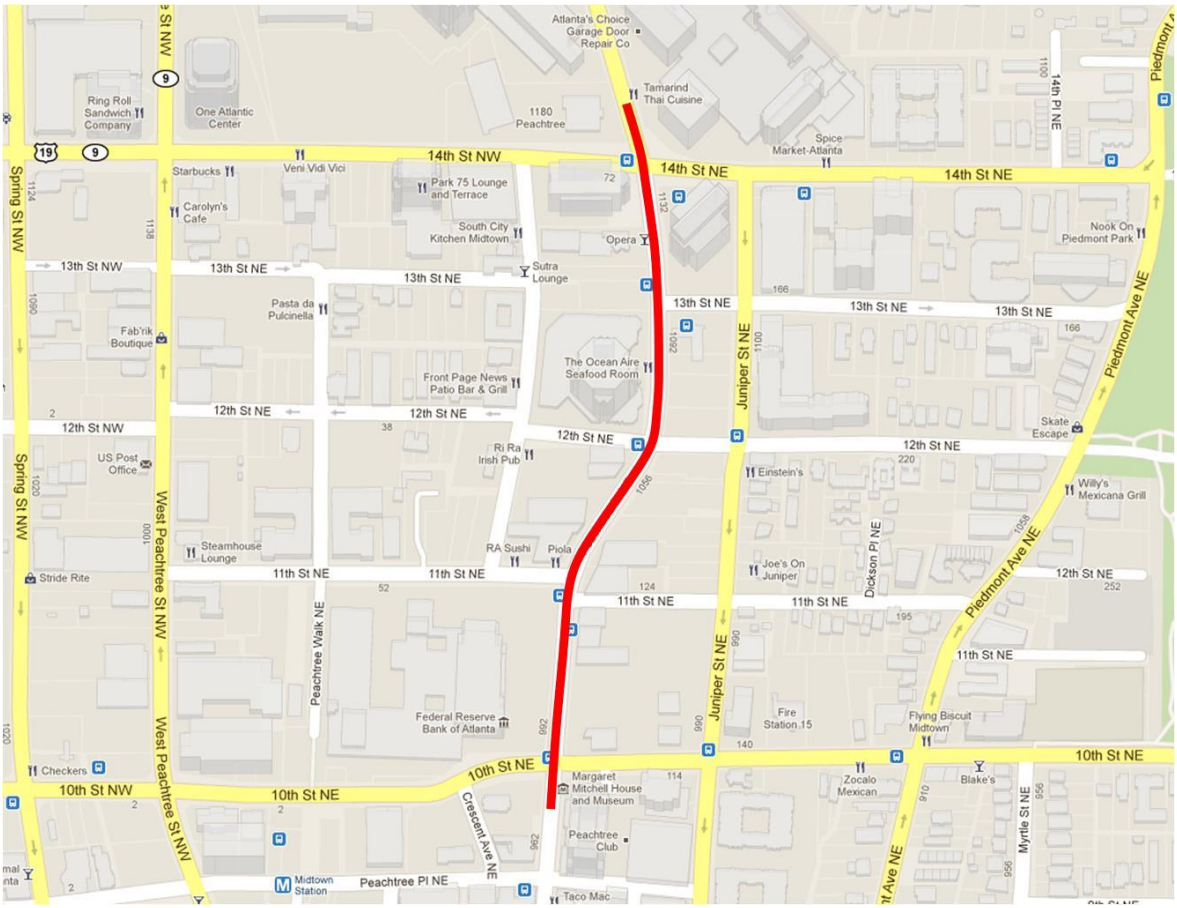


Figure 28 Peachtree Study Corridor [4]

The NGSIM Peachtree dataset comprises of trajectory data (with a resolution of a tenth of a second) for all vehicles traversing the corridor during the study period. Trajectory data was gathered on November 8, 2006, between 12:45PM and 1:00PM and 4:00PM and 4:15PM. In addition signal phase information at each intersection, origin-destination data (OD) for each vehicle, turning movement distribution at each intersection, and a series of other traffic related information were also collected. Video of the entire corridor is also available during data collection periods.

4.4.2 The Study

For the experiment, a detailed VISSIM model of the study area (Figure 29) was created. Roadway geometry was based on existing conditions at the time of the experiment and additional information such as vehicle volume, turning movement distribution, routing decisions and signal timing plans were added based on the NGSIM data set. Several verification iterations were completed to ensure that the model correctly represented the area being studied, as well as the traffic conditions during the study period. During this verification process issues related to the number of vehicles traversing the corridor and to signal timing plans were identified.

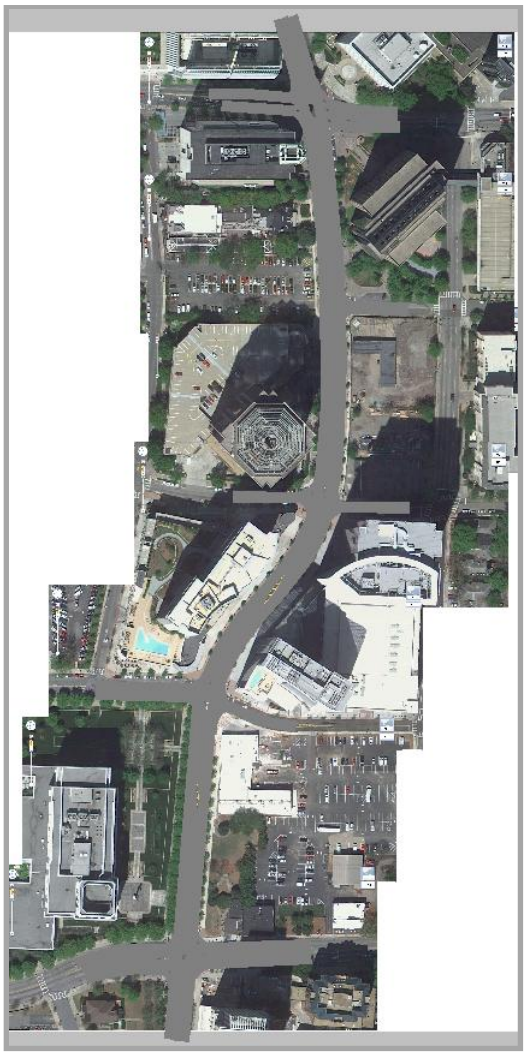


Figure 29 VISSIM model of Peachtree Study Corridor [4]

For instance, the number of vehicles, and subsequent turning movement distributions, were initially based on the summary reports produced by the NGSIM program. However, there were discrepancies between these summaries and counts extracted by hand from the videos. To address this issue, a software tool was developed to help record vehicular volume and turning movement counts from the videos. Errors identified in the NGSIM data were corrected, as best as possible, in the utilized VISSIM data set.

The NGSIM data also provided a direct observation of the signal indications. The observed signal indication did not appear to coincide with the provided signal timing controller data, likely indicating that the provided controller data was out of date. Thus, the signal indication data and engineering judgment was utilized to develop likely signal timing plans that would match the indication observations. The final simulation model is based on these plans, which includes offset observations. It is noted that during the observation periods (approximately 15 minutes) the offsets did appear to drift by a few seconds in the NGSIM data. To address this issue an average estimated offset is utilized.

4.4.2.1 Simulating Data Stream

To simulate streaming detector data, trajectory and OD information were used to create a VISSIM trip-chain file, which approximates the process of streaming detector data into the real-time simulation. A trip chain file is able to approximate a detector stream as each file's record consists of a time-stamp, indicating when a vehicle entered the network (i.e. crossed a boundary link detector), and a zone number indicating a vehicle's origin (i.e. the boundary detector crossed) and destination. This string of information is similar to that from a detector, except for a vehicle's destination. However, destination zones are often times approximated through historical turning movement.

With this method of approximating streaming detector data, a number of issues were encounter during the creation of the trip-chain file. The most important piece of information that is needed to create a trip-chain file is correct OD pairs. However after examination of the OD information given by the NGSIM program it was noticed that a number of OD pairs were potentially incorrect. For example, there were OD pairs that suggested an unusually large number of vehicles performed a u-turn maneuver. To verify

these maneuvers the OD distribution tables from NGSIM's Summary Reports and the corresponding videos were examined [44], [45].

Simultaneous examination of the distribution tables and videos revealed errors associated with assigned OD pairs. For instances, the u-turns were often left turns from the mainline to approaches leaving the network. In addition, there were assigned OD pairs that were not traversed by any vehicles during the study period. These errors largely occurred when the tracking software lost its handle on a vehicle that it had identified and began a new track for the same vehicle. To correct these issues engineering judgment was use to identify potentially erroneous OD pairs and necessary corrections were made by observing the vehicle on the video.

4.4.3 Preliminary Results and Analysis

Using the corrected NGSIM data final trip-chain files where created and used for the VISSIM data input. Ten replicate runs were conducted for the comparison between field and simulated performance measures. Similar to previous tests, travel time is the performance measure monitored and compared. Table 11 presents a summary of the simulated travel times from each of the 10 replicates and a summary of the field travel times. Note, data corresponding to the 12:45PM – 1:00PM period is referred to as the Noon period and 4:00PM – 4:15PM is referred to as Evening.

Table 11 NGSIM vs. VISSIM (VSM) Travel Time Results

		Noon - North		Noon - South		Evening - North		Evening - South	
		Avg (s)	Std Dev	Avg (s)	Std Dev	Avg (s)	Std Dev	Avg (s)	Std Dev
VSM Run #	1	120.0	31.7	102.0	20.5	108.1	28.8	100.3	29.0
	2	120.7	33.9	100.9	21.7	113.3	32.4	99.3	29.8
	3	117.4	34.5	98.4	22.9	111.5	30.4	94.6	27.6
	4	118.1	32.2	98.6	25.1	118.8	29.8	100.2	28.2
	5	116.0	31.5	96.6	22.6	116.7	34.0	103.1	29.5
	6	112.5	31.1	98.8	23.4	112.5	32.1	104.8	30.1
	7	113.7	32.2	96.4	24.6	114.2	32.2	102.4	29.8
	8	119.0	33.1	99.3	22.9	113.6	36.0	99.8	28.9
	9	116.5	31.8	100.0	20.4	108.7	33.6	104.2	31.5
	10	113.8	31.2	95.1	25.8	113.1	32.7	103.1	28.8
VSM Avg.		**116.8**	**32.3**	**98.6**	**23.0**	**113.1**	**32.2**	**101.2**	**29.3**
NGSIM		**131.5**	**36.7**	**106.6**	**17.1**	**140.4**	**35.4**	**133.9**	**29.6**
% Error		**11.2**	**11.8**	**7.5**	**34.1**	**19.5**	**9.0**	**24.4**	**0.9**

In the following discussion the referenced VISSIM results are the average of the 10 replicate runs. It is noted that there are some discrepancies between the simulated and field travel time estimates. A key difference is that VISSIM tends to under estimate field travel times. The smallest difference between VISSIM and field travel time is approximately eight seconds, occurring for the Noon-South time period. While the largest difference, 32.7 seconds, occurred for the Evening-South period. The simulation does a slightly better job estimating travel times for the noon period versus the evening. When comparing the standard deviations in Table 11, the values produced by VISSIM are similar to those from the field. This is observation is encouraging as it indicates that VISSIM's approximation of the variation travel time estimates is rather similar to that of the field. With dissimilar means and "similar" standard deviations it is anticipated that the observed discrepancies may be addressed through a more rigorous calibration effort. Density plots were examined to further corroborate this hypothesis, see Figure 30.

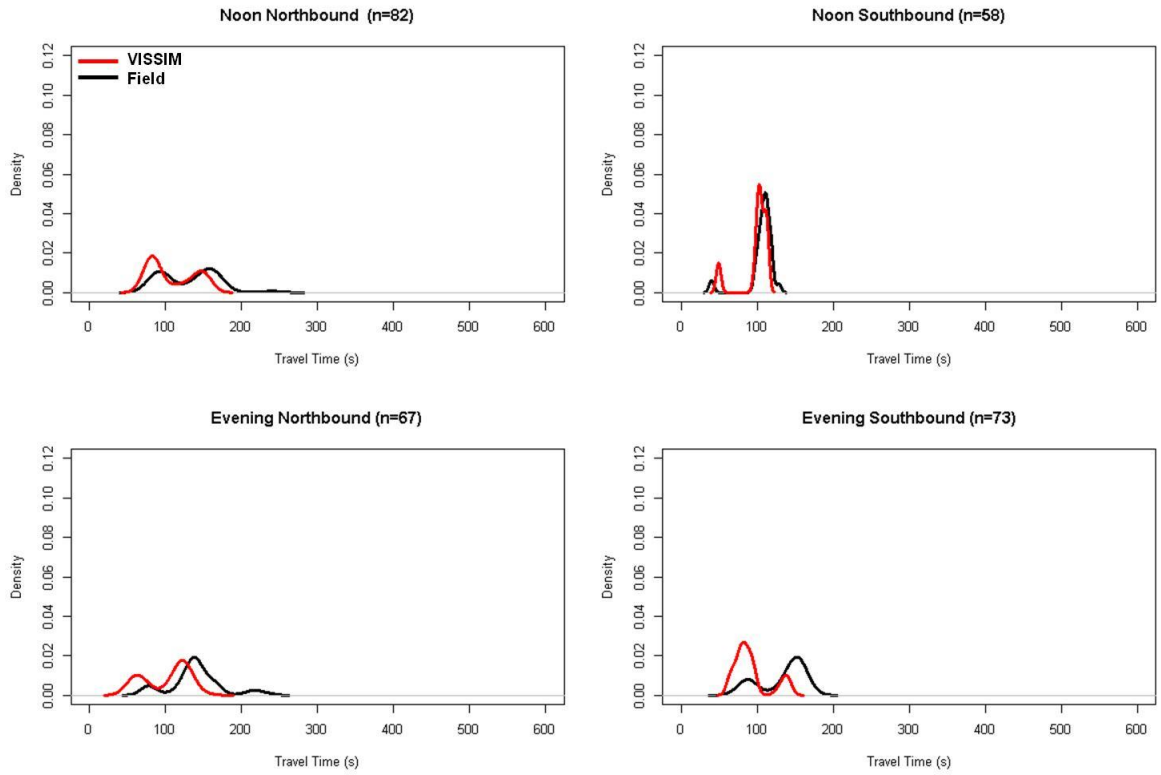


Figure 30 Density Plots of Field vs. VISSIM (single run) Travel Times

The density plots of the simulated travel times generally capture the bi-modal or tri-modal form of the field travel times. The differences between the plots tend to be a shifting of the centroid of the modes or proportionality between the different modes. However, in all cases the general form of the distribution is reflected, a very encouraging finding, likely indicating many of the differences can be addressed in calibration.

Finally, in addition to the travel time distribution plots the Time-Space-Diagrams for the field and simulated data were generated. Distinct discrepancies in driver behavior were observed as illustrated in Figure 31 and Figure 32.

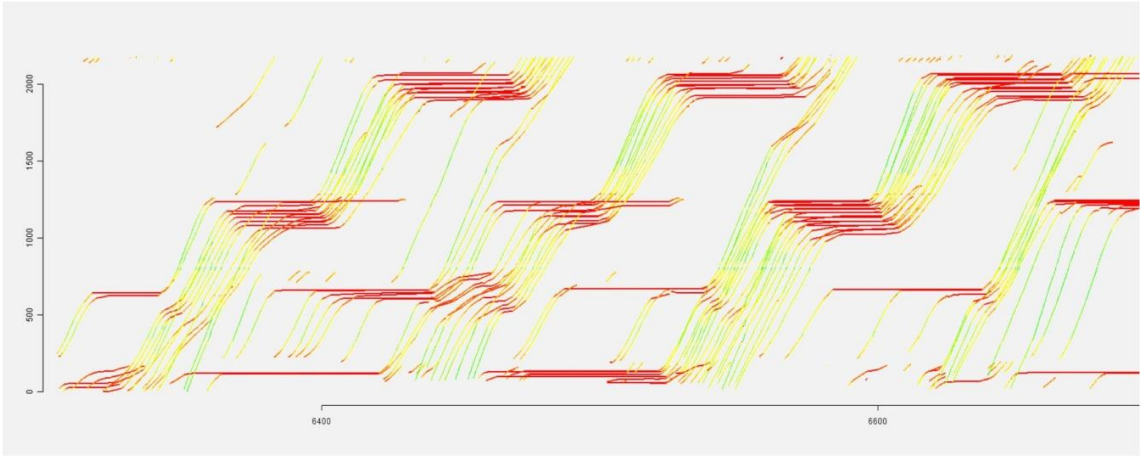


Figure 31 Northbound Real-World Time-Space Diagram

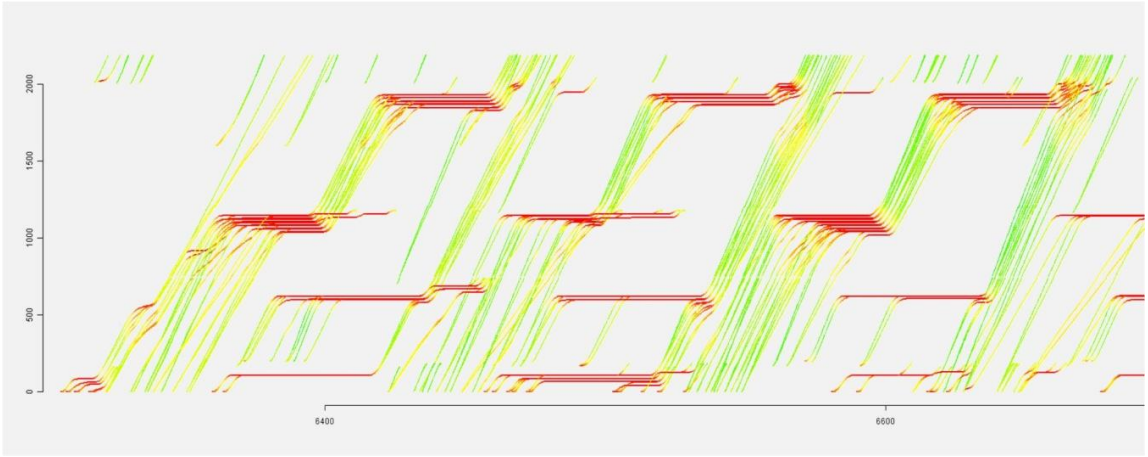


Figure 32 Northbound VISSIM Time-Space Diagram

In the above figures each line represents the trajectory of a vehicle as it traverses the length of the study corridor with respect to time. The colors along each trajectory represent current vehicle speed relative to the maximum speed along the corridor. Red represents speeds of approximately zero mph, green represents speeds of approximately 35 miles-per-hour, and shades of each color represent speeds between 0 and 35 miles-per-hour. In comparing field and simulated trajectories it is apparent the simulated traffic is

less variable (i.e. the traffic flow is “smoother”), with less interaction between vehicles. One of the more recognizable differences is that simulated vehicles tend to achieve their desired speed more quickly and maintain that speed for longer periods. A likely reason for this difference is that simulated drivers are being modeled with more aggressive tendencies than their field counterparts and less variability in aggressiveness across drivers. As a result of this more aggressive driving by simulated vehicles they will tend to traverse the corridor in less time versus vehicles in the field, and may clear an intersection during the green or amber phase while their field counterpart may not make through that intersection at that point in time. Such scenarios are supported by the travel time measurements that are presented in Table 11 as VISSIM tends to underestimate real-world travel times.

Given the above results for the NGSIM pseudo field experiment and the insights afforded by the time space diagrams it is anticipated that more accurate estimates of field travel times may be achieved through an advanced calibration procedure. This procedure will involve a Monte Carlo parameter selection method which determines the most effective parameters to calibrate a VISSIM simulation model. Chapter 5 will present this proposed calibration procedure in detail.

4.5 Experimentation and Evaluation – Summary

Simulated measures did not perfectly reflect those from the field. However, the relative accuracy between the two is able to equip travelers and facility managers with information that encourages more informed usage and operating decisions. The preceding experiments have also informed future improvements to the current methodology. A few of the measures will be presented in Chapter 6, Future Research .

CHAPTER 5

MODEL CALIBRATION

The results from the NGSIM corridor model tests highlight the feasibility of using a data-driven simulation approach to effectively estimate arterial performance measures in real-time. This pseudo-field test also demonstrated that a well-calibrated model might lead to more accurate performance measure estimates.

This chapter presents an advanced calibration procedure that utilizes candidate sets of parameter values that were evaluated using a series of selection rules to determine which parameters set(s) result in a well calibrated model. In addition, specific guiding principles for assigning initial values to calibration parameters are also discussed. In both cases, ten potential calibration parameters of the previously discussed NGSIM model are used as a test case. This calibration process seeks to inform the development of more robust calibration procedures that will subsequently lead to more accurate performance measure estimations, which is imperative for future real-time (and off-line) simulation efforts.

5.1 Selecting Effective VISSIM Calibration Parameters

The following summarizes a sensitivity analysis process used to select parameters for calibrating VISSIM arterial models. For additional details regarding this process, readers are encouraged to consult Miller (2009) [46] and Miller et al. (2012) [47].

The sensitivity analysis process consists of four steps: 1) initial parameter selection, 2) performance measure selection, 3) Monte Carlo simulations, and 4) parameter elimination. Steps 3 and 4 are repeated until a pre-defined stopping condition is satisfied. Miller [46] demonstrated that approximately half of VISSIM's 50 calibration

parameters could be eliminated in the first step due to the facility type being modeled (i.e. arterial versus freeway). In step two, travel time was selected as the performance measure to ascertain the effect of the calibration parameters on the model. Step three in Miller's study involved a Monte Carlo experiment that generated a series of simulation runs used to isolate the sensitivity of the model performance (i.e. travel time) to the parameter values. One thousand parameters sets were created in a Monte Carlo fashion. The value for each of these parameters was limited by its reasonableness range, determined as part of step 1. These parameter sets were then used as inputs to generate 1000 unique VISSIM simulation runs where only these parameter values varied between runs. From each simulation run travel time measures were extracted and analyzed. Step four compared travel time measures with the values for each parameter and eliminated parameters whose values had little or no effect on travel time. Default values were assigned to the eliminated parameters and step three was repeated until no parameters could be eliminated due to a lack of influence on travel times. Figure 33 illustrates the execution of these four steps. [46], [47]

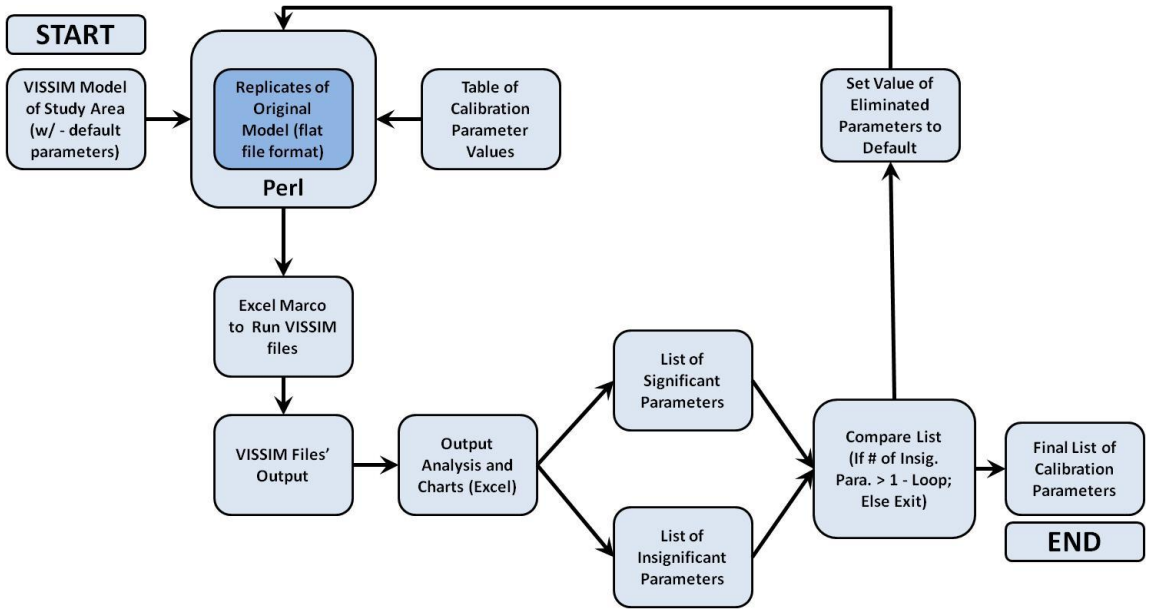


Figure 33 Four-Step Sensitivity Procedure to Select Effective Parameters [47]

This 4-step sensitivity analysis procedure was carried out using a VISSIM model of Cobb Parkway (U.S. Highway 41) – a 1.4 mile, 11 intersection arterial, in Cobb County GA. From a set of 22 calibration parameters relevant to arterial simulation models (Table 12), nine were selected as influential parameters (Table 13). These nine parameters were therefore recommended as parameters that should be considered for calibration of VISSIM arterial models

Table 12 Relevant Arterial Calibration Parameters and Their Ranges [46]

#	Parameter	Initial Range	Final Range
1	Desired Speed Distribution Range	±0.0-10.0 mph	±0.5-10.0 mph
2	Look-ahead distance min	0-900 ft	0-300 ft
3	Look-ahead distance max	500-1000 ft	500-1200 ft
4	Number of observed vehicles	2-8	
5	Average standstill distance, ax	0.0-20.0 ft	2.0-8.0 ft
6	Additive part of safety distance, bx_{add}	0.0-8.0	0.0-3.0
7	Multiplicative part of safety distance, bx_{mult}	0.0-8.0	0.0-3.0
8	Maximum Deceleration (own)	-20.0 - -3.0	
9	Maximum Deceleration (trailing)	-20.0 - -3.0	
10	Accepted Deceleration (own)	-6.0 - -0.33	
11	Accepted Deceleration (trailing)	-6.0 - -0.33	
12	Reduction rate (own)	50-300	50-200
13	Reduction rate (trailing)	50-300	50-200
14	Waiting time before diffusion	20–80 sec	40–80 sec
15	Min. headway (front/rear)	1.64-25.00 ft	
16	Safety distance reduction factor	0.0-1.0	
17	Max. deceleration for cooperative braking	-35.0 - -3.0	
18	Reduction factor lane change before a signal	0.3-0.9	
19	Start upstream of stop line	200–600 ft	
20	End downstream of stop line	200–600 ft	
21	Emergency stop distance	6.56–30.0 ft	
22	Lane change distance	300–1000 ft	500–1000 ft

Table 13 Final Set of Effective Calibration Parameters [46]

#	Parameter	Initial Range	Final Range
5	Average standstill distance, ax	0.0-20.0 ft	2.0-8.0 ft
6	Additive part of safety distance, bx_{add}	0.0-8.0	0.0-3.0
7	Multiplicative part of safety distance, bx_{mult}	0.0-8.0	0.0-3.0
8	Maximum Deceleration (own)	-20.0 - -3.0	
9	Maximum Deceleration (trailing)	-20.0 - -3.0	
15	Min. headway (front/rear)	1.64-25.00 ft	
16	Safety distance reduction factor	0.0-1.0	
17	Max. deceleration for cooperative braking	-35.0 - -3.0	
22	Lane change distance	300–1000 ft	500–1000 ft

5.1.1 Desired Speed Range Parameter

In the Miller [46] effort elimination of the desired speed range parameter from the set of effective calibration parameters seemed counterintuitive. This parameter dictates the variability in a drivers' desired speed and it was anticipated that travel time would be sensitive to this parameter. However, this was not the case in the Miller [46] experiment. To explore this result an experiment was conducted using the NGSIM VISSIM model. In the experiments carried out by Miller [46] the desired speed range parameter varied from zero to ten miles-per-hour (implying desired speed could vary up to +/- 10 mph). For this new experiment, the desired speed range parameter ranged from zero to 20 mph. Approximately 300 simulation models were created, based on the NGSIM corridor model. Each of these models had the same values for each calibration parameter, except for the desired speed range parameter. Travel time measurements were extracted after each replicate run. Figure 34 presents a scatter plot of desired speed range versus average travel time for each segment. Each point presents a segment's average travel time and its corresponding desire speed range, and the red horizontal line represents the average travel time from the field.

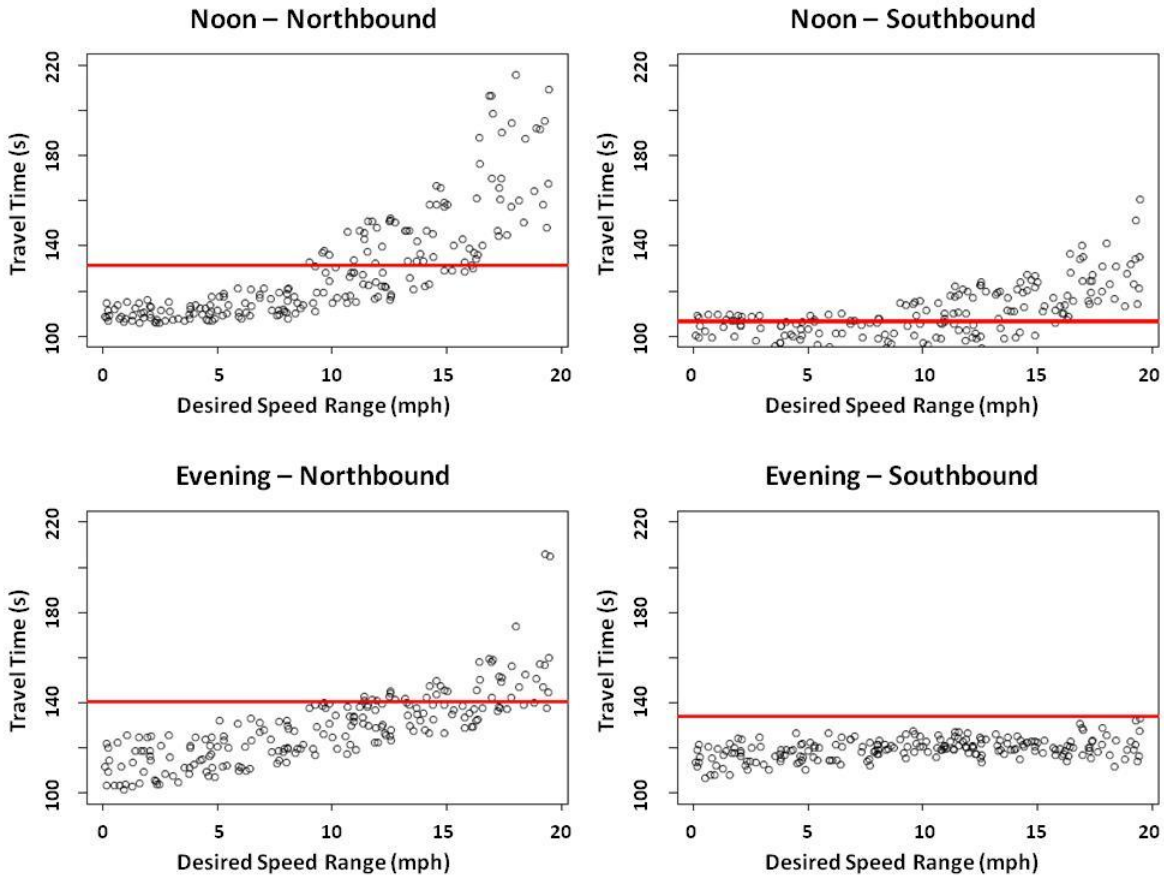


Figure 34 Average Travel-Time versus Desired Speed Range

In Figure 34 there is little change in travel time when the desired speeds range value is between zero and 10 mph, although a trend is apparent for the evening northbound simulations. This result is consistent with the decision to eliminate the desired speed range parameter from the list of effective calibration parameters. However, as the desired speed range increases beyond 10 mph the average travel time estimates begin to change. Since it is not known from this experiment if other models would have a different range sensitivity (for instance the sensitivity to range may be influenced by signal coordination parameters, with potentially smaller ranges having an influence) the desired speed range parameter was included as a parameter requiring calibration for the purposes of the study.

5.1.1.1 The Width of the Desired Speed Range Parameter

In this study, the desired speed range varied from zero to 20 mph. Such a wide range has the potential to produce unrealistic desired speeds. For instance, if the speed limit on an arterial is between 25 mph or 35 mph, this range will produce minimum desired speeds between 5 mph and 15 mph. Although, such low speeds are possible, it is unlikely that they accurately reflect drivers' desired speeds, especially along this corridor. To explore the validity of this assumption, desired speed estimates from the NGSIM dataset were examined.

To estimate drivers' desired speeds, mid-block speeds were extracted from five locations along the Peachtree corridor, see Figure 35. Speeds at these mid-block locations are believed to be representative of a driver's desired speed as the effect of traffic signals is at a minimum and low traffic volumes along this roadway during the study period should allow the drivers to reach their desired speeds. To determine desired speed, vehicle trajectory data were extracted from 30-50 foot segments at these mid-block locations for analysis.

Figure 35 NGSIM Peachtree Corridor with Mid-Block Segments

NGSIM's vehicle trajectory dataset include: vehicle ID, speed, OD pair, (x, y) coordinates, heading, and lane number, amongst other relevant vehicle data. These data were collected at a resolution of 10 Hz. Trajectories of only vehicles that traversed the entire length of the corridor were considered for this analysis to avoid potential bias generated by a vehicle seeking a turn. Each of these vehicles had at least 5 to 10 data points belonging to each of the selected mid-block zones. For a given vehicle, two sets of measurements were recorded from each zone. These measurements were the average and maximum speed for that zone. The two estimates are noted as Avg_Avg, for the average of the average speeds in each zone, and Avg_Max, for the average of the maximum speeds in each zone. Figure 37 and Figure 37 present the histograms of these two estimates for the vehicle traveling during the Noon period. Figure 38 and Figure 39 present the same results for those traveling in the Evening period. Table 14 then provides the summary statistics for both periods and both sets of estimates.

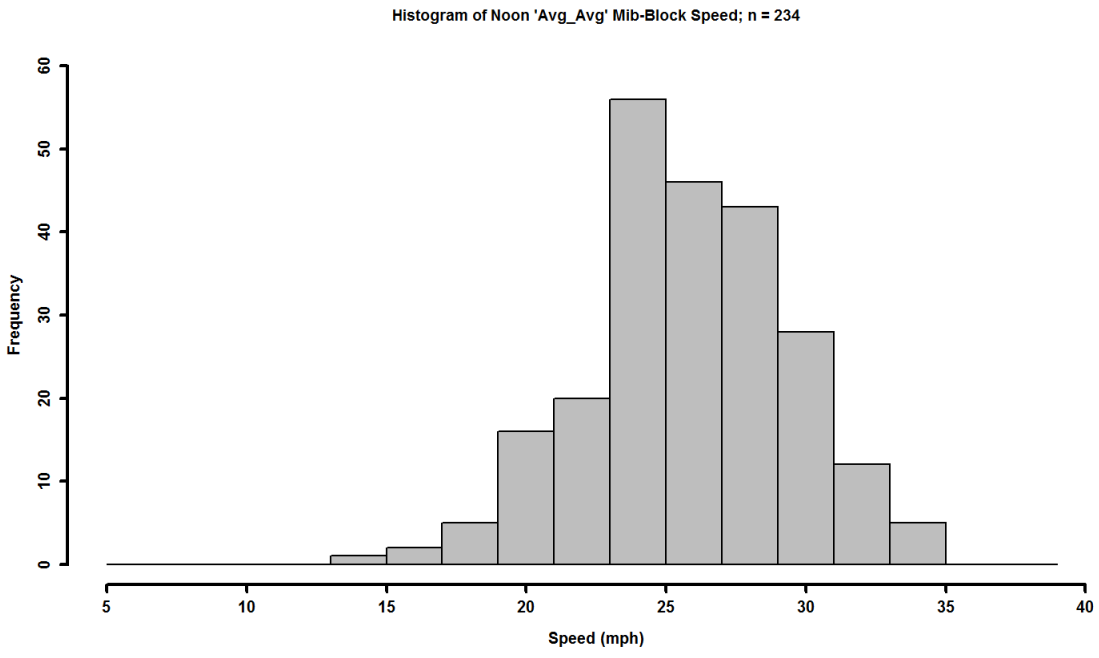


Figure 36 Histogram of Estimated Desired Speeds (Avg_Avg) – Noon

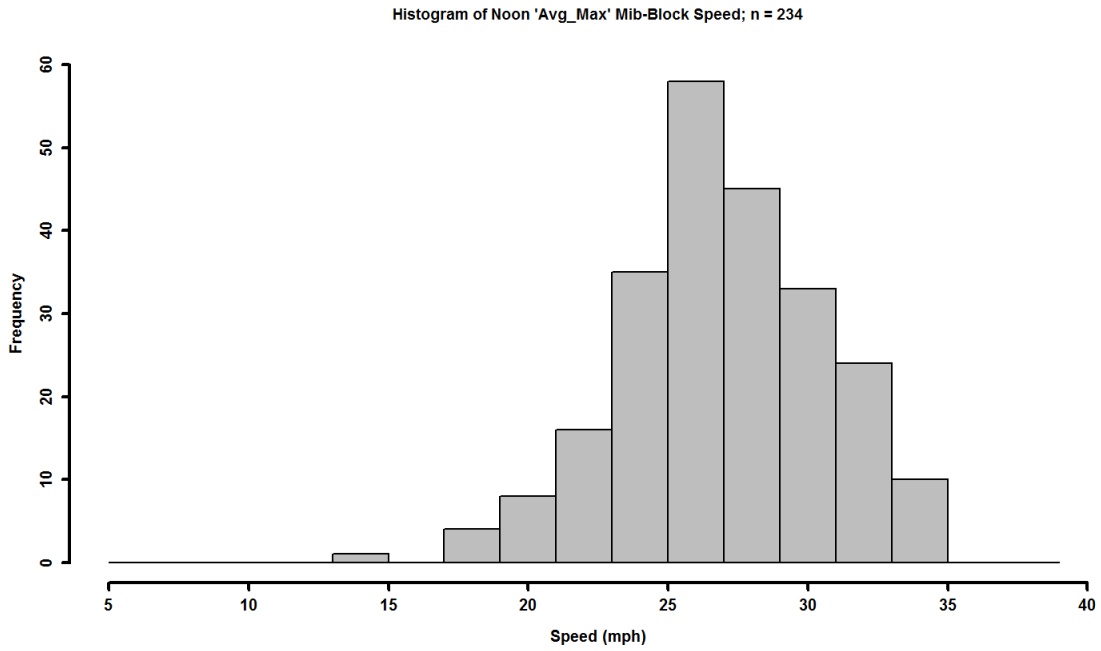


Figure 37 Histogram of Estimated Desired Speeds (Avg_Max) – Noon

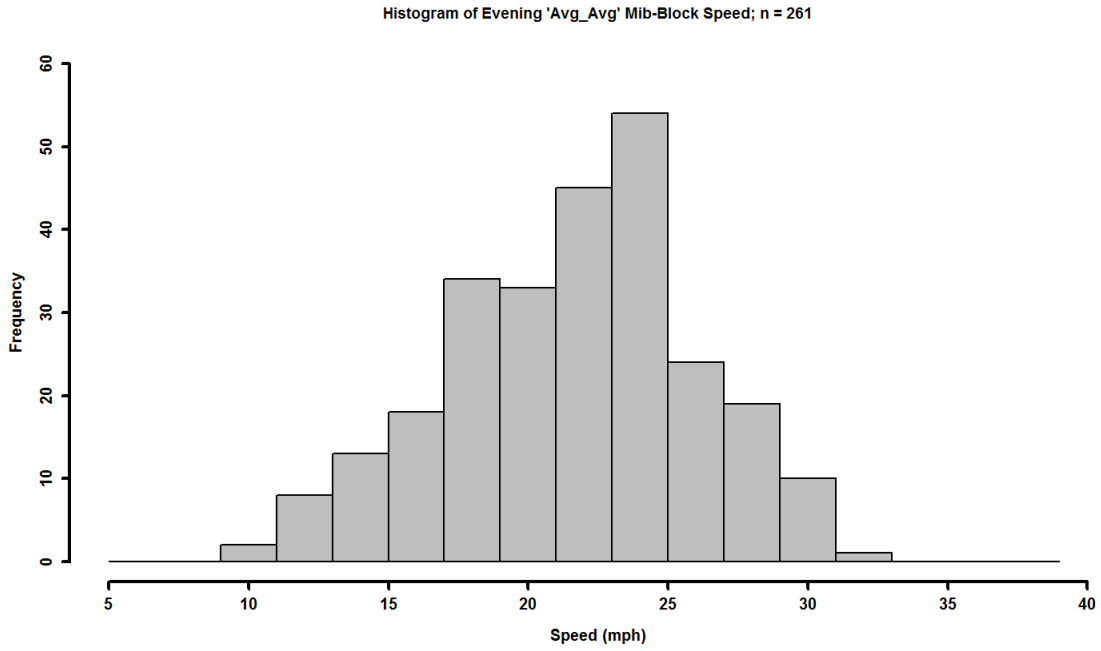


Figure 38 Histogram of Estimated Desired Speeds (Avg_Avg) – Evening

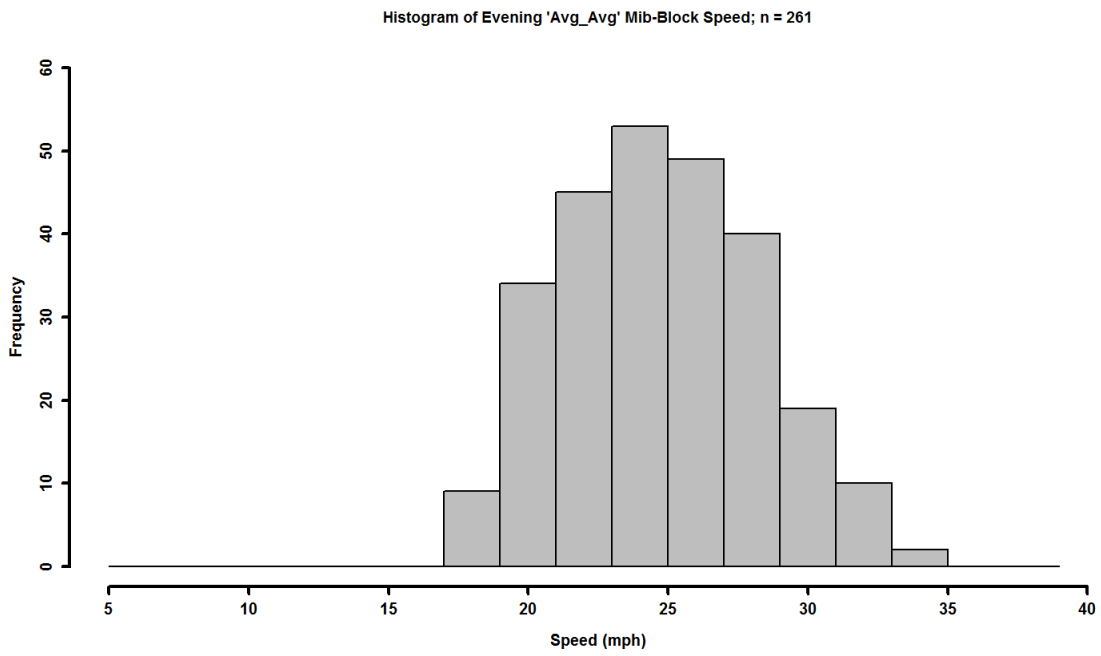


Figure 39 Histogram of Estimated Desired Speeds (Avg_Avg) – Evening

Table 14 Summary Statistic of Estimated Desire Speeds

Statistic	Noon Period		Evening Period	
	Avg_Avg (mph)	Avg_Max (mph)	Avg_Avg (mph)	Avg_Max (mph)
Min	14	15	10	18
Mean	26	27	22	25
85th %	30	31	26	29
Max	35	35	32	33
Standard Dev.	4	4	4	3

Given the above summary statistics of field desired speed estimates, maintaining a desired speed range between 0 and 20 mph does potentially result in unreasonable values. A more reasonable range for this parameter is between 0 and 10 mph. This research effort however proceeded with the more conservative range of 0 to 20 mph to ensure that all desired speed estimates, especially the minimums, which range from 10 to 18 mph, were thoroughly examined as a part of the calibration process. However, after conducting the

necessary calibration exercises, the modeler should examine the mean desired speed and desired speed range to ensure that the selected values of these measures in the calibrated model are reasonable.

5.2 NGISM Model Calibration

Based on the above procedure, the ten final parameters were employed to calibrate the NGSIM corridor model discussed in Section 4.4. To assist in the calibration process a Monte Carlo inspired approach was adapted from Miller (2009) [46] to create candidate parameter sets that sufficiently represented the sample space for each parameter and the various combinations. The Monte Carlo approach produced 1000 unique parameter sets that were simulated in the NSGIM model. From these model runs, travel time and saturation flow measures were extracted. These measures were then analyzed to determine which combination of parameter values most closely reproduced the NGSIM results.

A *batch-means* [48] inspired method was employed to extract data during this calibration process. The base simulation model, in which all parameter sets were tested, consisted of three "batches" with a total simulation period that far exceeded the sum of all study periods. These three batches represented the 15- minute Noon and 15- minute Evening NGSIM study periods, and an hour-long saturation flow study period. The primary differences amongst these batches are signal timing plans, routing decisions and vehicle volumes. An approximate 4-hour simulation period was used to ensure independence between batches. The first study period was simulated between 0-5400 seconds, the second – 6300-7200 seconds, and the third 10000-14000 seconds. Each

simulated period began with a transient period during which the simulation approaches steady state. No data were extracted during this period. To ensure independence between simulation periods (batches), there was a transition period between each. This transition period provided an opportunity for all the remaining vehicles, at the end of a simulation period, to exit the model before the following period commences. Having an empty model before the next simulation period begins is how independence between the simulation periods (batches) was maintained. This method efficiently executes multiple simulation studies, that are sufficiently similar, while maintaining the integrity of each study's analysis.

Figure 40 - Figure 43 presents travel time density plots for each NGSIM study period and travel direction. Each of the 1000 VISSIM (VSM) simulation runs, each containing the described three-batches, produced these density plots. This set of results will be used in conjunction with saturation flow measures to aid in the NGISM model calibration process.

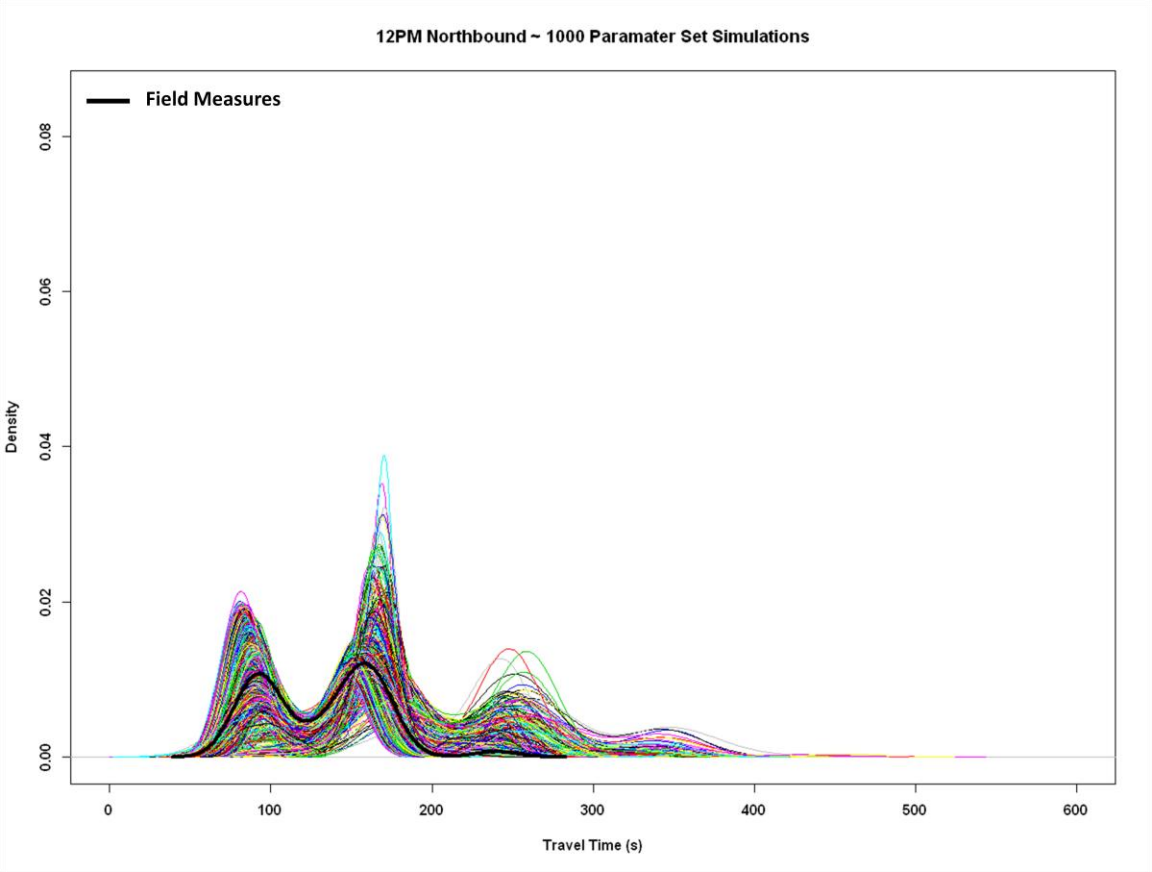


Figure 40 Noon Travel Time Density Plots – Northbound

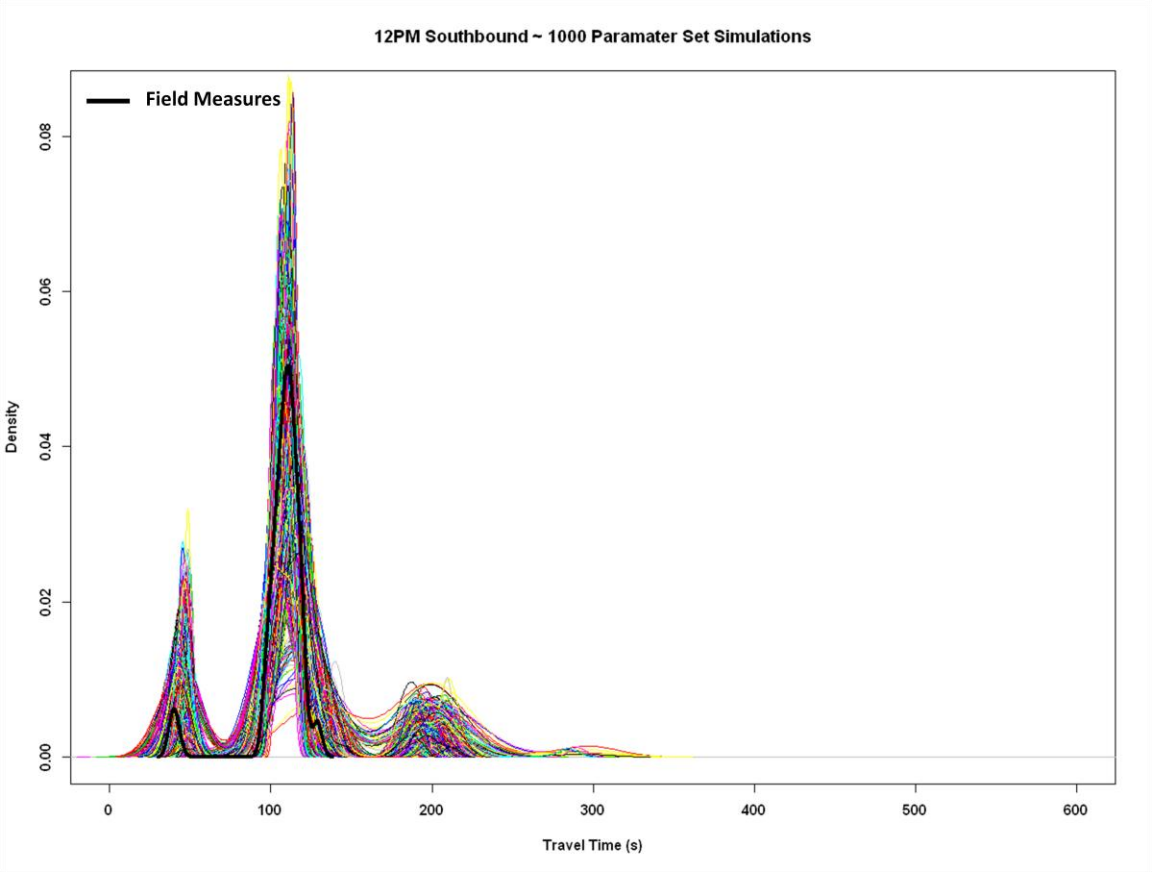


Figure 41 Noon Travel Time Density Plots – Southbound

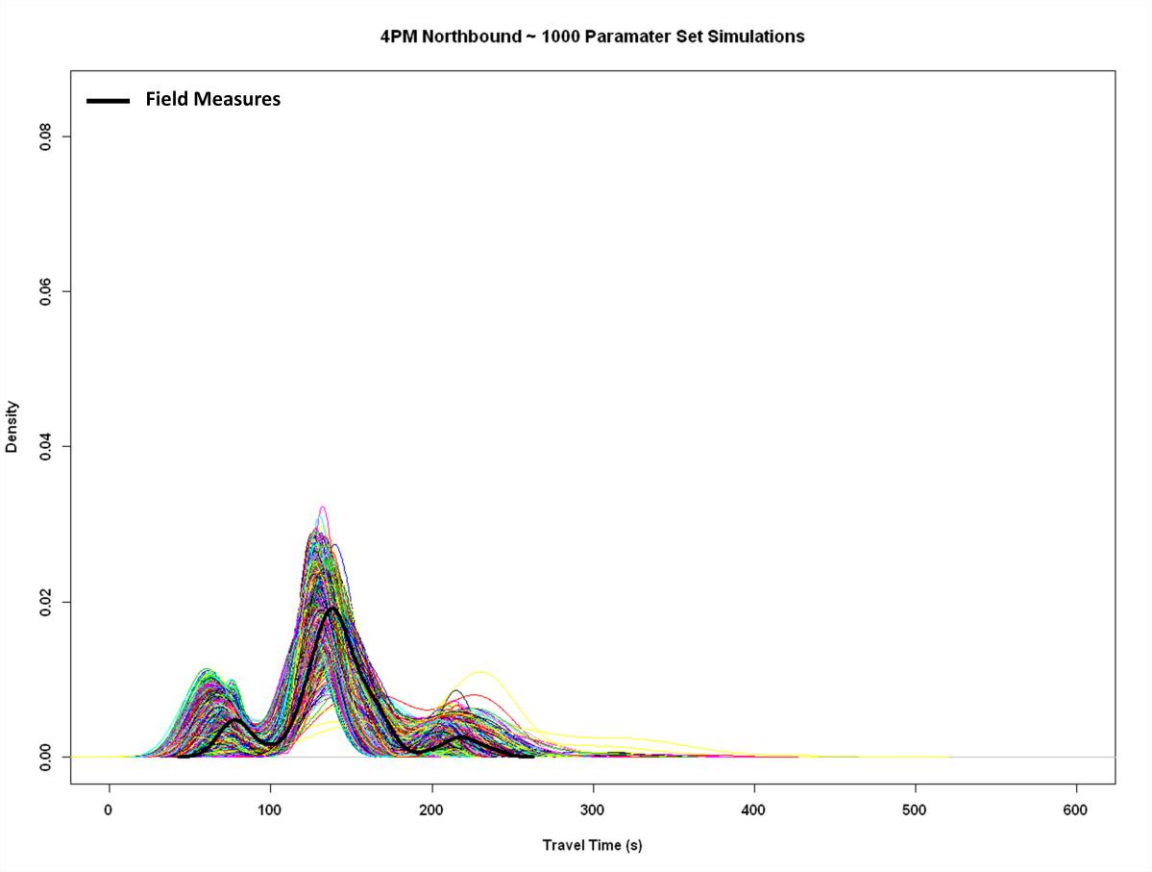


Figure 42 Evening Travel Time Density Plots – Northbound

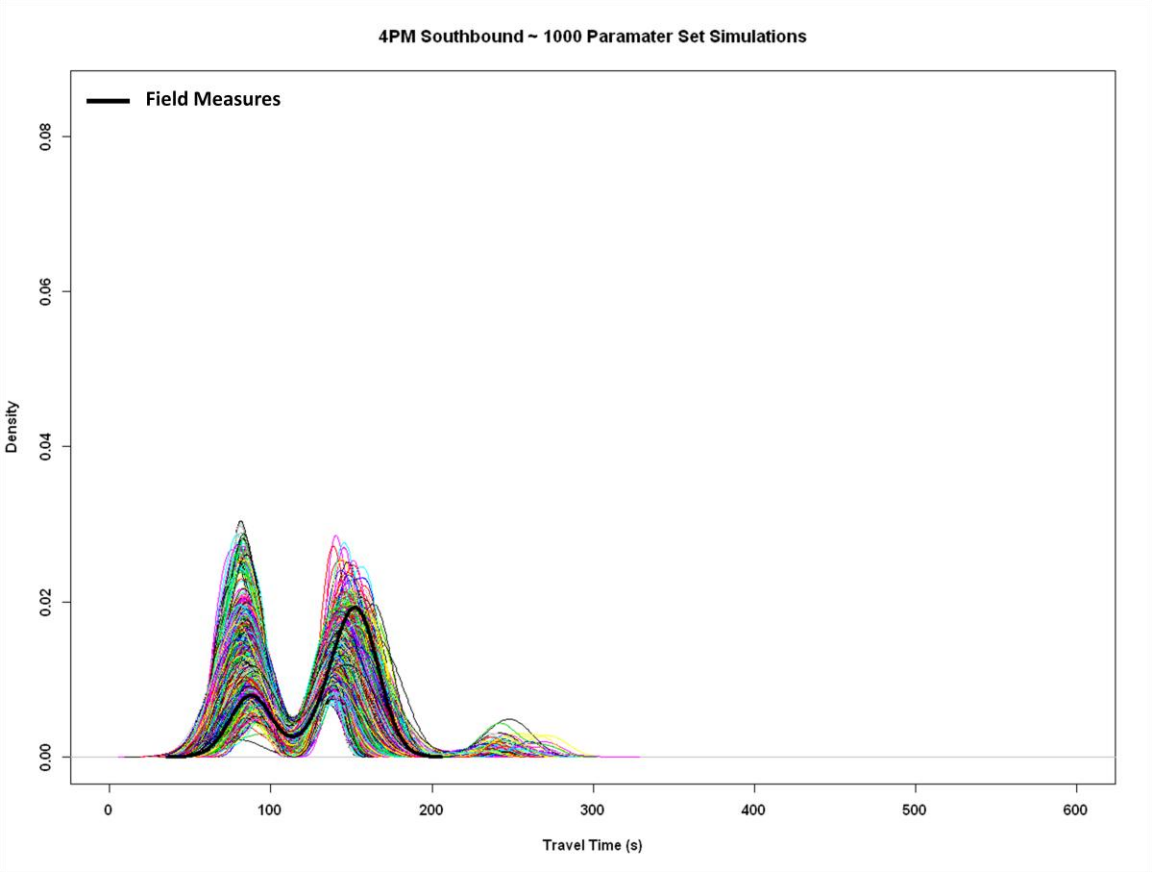


Figure 43 Noon Travel Time Density Plots – Southbound

From Figure 40 - Figure 43, there appears to be a number of parameter sets that yielded reasonable estimates of field travel times, although parameters set were also generated that significantly differed from the field. The following sections will detail the development of a robust procedure to select calibrated parameter sets from the candidate sets. In addition, the following sections seek to aid the simulation community in assigning values to the calibration parameters. It is hoped that such information will inform and increase the efficiency of the calibration process for a VISSIM arterial model.

5.2.1 Selecting a Calibrated Parameter Set

One of the difficulties in selecting a parameter set that result in a calibrated simulation model is establishing suitable criteria that indicate when a model is calibrated. The criteria established by the traffic simulation community tend to be subjective due to their dependency on what is being modeled and the intentions for which the model was created [49]. To date, the criteria to determine whether or not a model is calibrated typically involves parametric, first moment statistical comparisons of field and simulated performance measures. Park and Schneeberger (2003) [50] used the results from the t-test to compare simulated and field travel time means as the criterion to determine an adequately calibrated model. Park and Won (2006) [51] developed a criterion which deemed a model as calibrated when the model's travel time distribution "include the entire field-measured values" (p. 50).

Traffic simulation models are often created with the intention of evaluating alternative scenarios to improve current conditions or to appropriately plan for future changes in demand on the transportation network. Examining mean performance measures may be a sufficient calibration criterion to provide a general analysis of future scenarios; however, it is questionable if calibration on mean values is sufficient. For simulated estimates, the more the higher-order moments differ from the field distribution, the less confidence a user has in the ability of the simulation to accurately reflect alternative scenarios. For example, in Park and Schneeberger (2003) [50] the calibrated model accurately estimates mean travel time. However, when considering the form of the travel time distribution, discrepancies were clear. See Figure 44. In this instance, the field data is significantly bi-modal, a characteristic missed in the "calibrated" model.

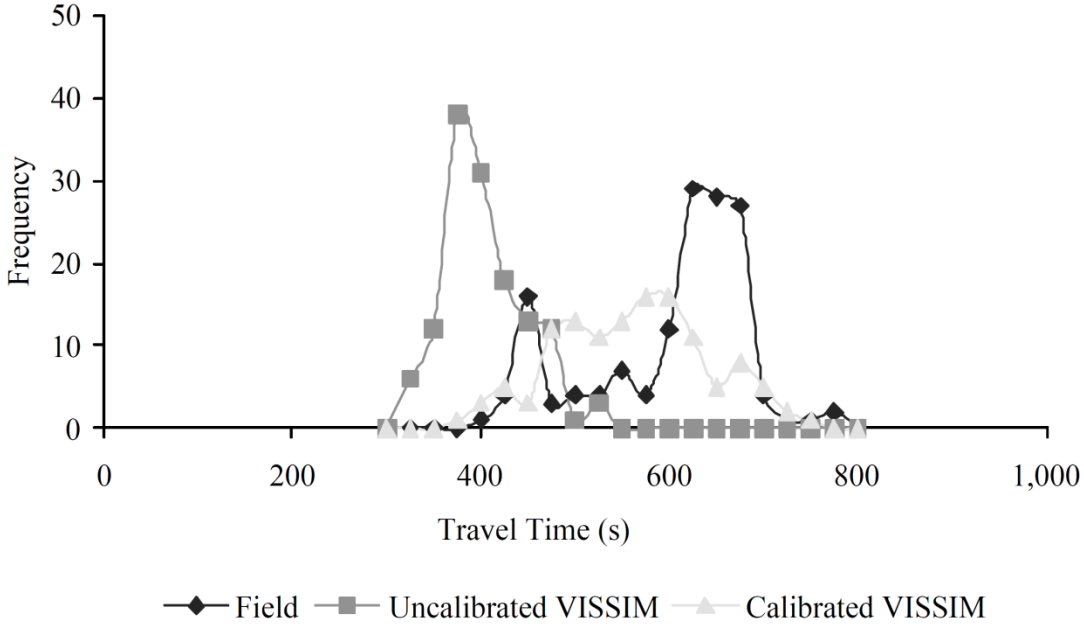


Figure 44 Comparing Field Data, Uncalibrated VISSIM, and Calibrated VISSIM [50]

For real-time simulation, more robust calibration criteria are clearly needed. The primary impetus for having more involved calibration criteria is to increase the quality and accuracy of information that is desired from some of today's more involved simulation models. These models seek to communicate both facility level (e.g. travel time over a corridor) and time or vehicle specific (e.g. predicated arrival time of a vehicle at an intersection) information to travelers and facility managers. This information will in turn arm these consumers with the knowledge to facilitate improved network usage and management. The accuracy of the information provided is therefore paramount (especially on the microscopic scale) hence the need for more robust calibration criteria to dictate when a model is adequately calibrated.

This research effort formulates a two-part criteria process to select calibrated parameter sets. The first part compares field and simulated saturation flow rates to ensure that the models produce reasonable estimates. This is an imperative step as it is possible for models to produce accurate estimates of performance measures, such as travel time, while over or under estimating saturation flow rates. Failure to adequately model saturation flow could prove significant. For example, by overestimating saturation flow a simulation will overestimate capacity. In a scenario analysis where base volumes are increased to a higher level to represent build conditions it is possible they could exceed field capacity (implying significant congestion) but the simulation would continue to show uncongested operations. As a result, the simulation would no longer reflect field conditions.

The second part of this process involves the statistical evaluation of the mean and the distribution of the performance measures being studied, travel time in this case. By evaluating travel time means and distributions, the calibrated model will consider reflecting both corridor level and individual vehicle level traffic information, which is necessary for real-time simulation

5.2.1.1 Startup and Saturation Flow Criterion

Establishing a saturation flow criteria not only facilitates greater confidence in the results from an calibrated model but it also provides some level of protection from the potential dangers in the application and implementation of the calibrated model to alternative conditions.

Acceptable ranges for saturation flow were established based on appropriate field measurements. Videos of the NGSIM corridor operation were observed and saturation

headways were collected. Saturation headway "is the amount of time that a vehicle in the stopped queue takes to pass through a signalized intersection on the green signal, assuming that there is a continuous queue of vehicles moving through the intersection" [52], p. 7-8. To put it another way saturation headway is "the constant headway achieved is referred to as the saturation headway, as it is the average headway that can be achieved by a saturated, stable moving queue of vehicles passing through the signal" [53] p. 143. Headway measurements of the first four vehicles are not considered when estimating saturation headway. However, these headway measurements were included (referred to as startup flow) in this effort to allow for a calibration of parameters related to queue start-up as well as saturation flow, providing a better estimate of capacity. [53]

Headway measurements were collected and arranged in two groups. One group contained measurements from the second vehicle in the queue to fifth vehicle (startup flow estimate), while the next group contained measurements from the sixth to the ninth vehicle (saturation flow estimate). These headway measurements were then averaged for each cycle and converted to startup (s_{su}) and saturation (s) flows via Equation 1, [53]. The startup and saturation flow rates (per cycle) were then used to create density plots of the observed flow rates.

$$s_{su} = \frac{3600}{h_{su}} \quad and \quad s = \frac{3600}{h} \qquad \textbf{(1)}$$

where: s_{su} = startup flow (vehicles/hour)

h_{su} = startup headway (sec)

s = saturation flow (vehicles/hour)

h = saturation headway (sec)

Headway data was collected at the intersection of Tenth and Peachtree Streets. A screen line was placed in Lane 2, at the stop bar of the northbound Peachtree St. approach. Figure 45 presents the data collection location and a screenshot from the video that was recorded during the NGSIM study. Manual processing of the recorded video was assisted by using an in-house video processing software that recorded timestamps and frame IDs. For these data, when a vehicle crossed the screen line (manually identified) the user would press a computer key to record the timestamp and frame number. Timestamps were only recoded for the vehicles that were in the queue as the queue began to discharge. Field headway data and the corresponding estimates of startup and saturation flow are presented in Table 15.

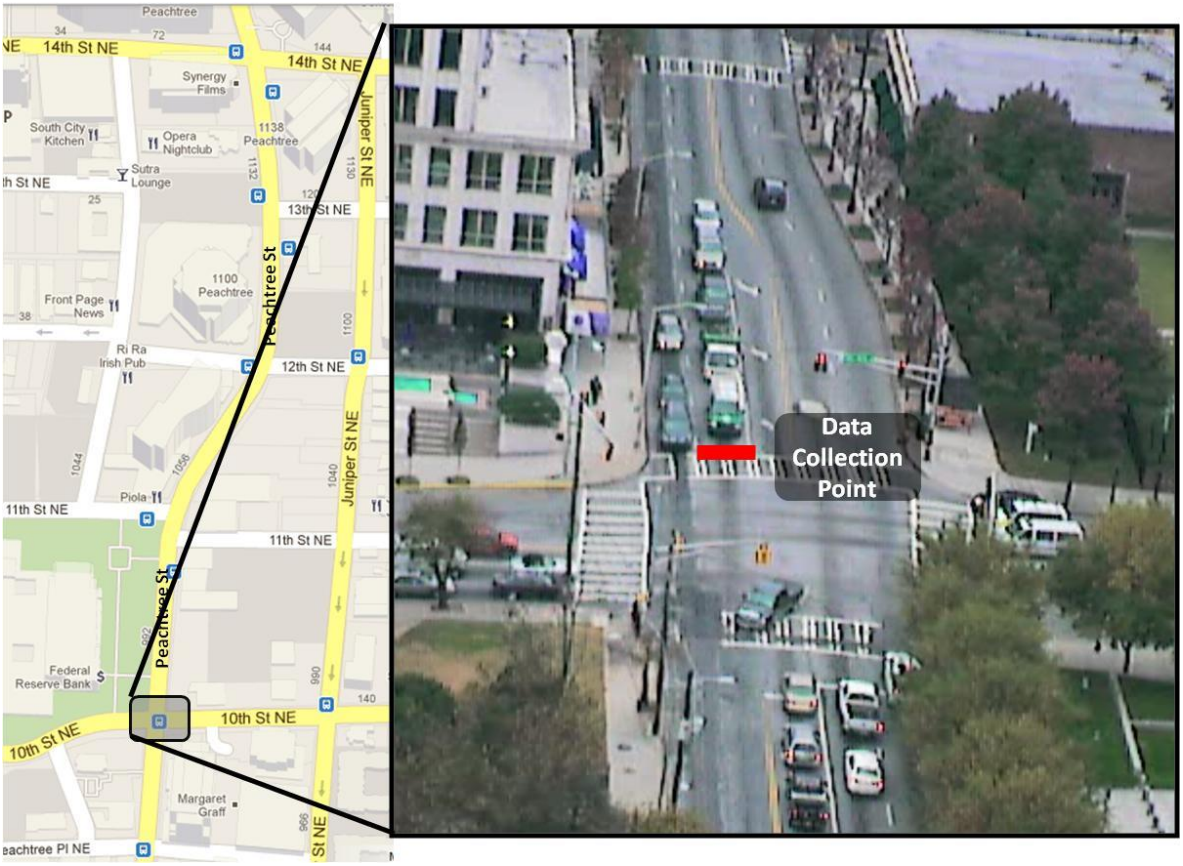


Figure 45 Location of Headway Field Data Collection

Table 15 Field Data – Headways, Startup Flow, and Saturation Flow

	Vehicle Position									Startup Flow	Saturation Flow
	2	3	4	5	6	7	8	9	10		
12 PM											
Cycle Number											
1	2.4	2.8	2.3	2.1						1498.6	-
2	3.6	3.4	2.3	2.0	2.6	1.7	1.5			1273.0	1883.2
3	2.8	2.3	2.5	2.2	2.1					1467.4	1694.1
4	3.1	2.3	2.0	2.6	3.1	2.8				1442.3	1228.9
5	3.7	2.5	2.2	1.9	1.9	2.1				1396.4	1806.8
6	2.7	2.2	1.9	2.6	1.7	2.7				1518.3	1651.8
7	2.9	2.9	1.7	2.3	2.9	3.3	1.8	2.6	2.2	1479.2	1361.3
8	2.1	2.7	2.1	3.0						1455.9	-
9	2.8	3.0	1.9	2.1						1472.2	-
10	2.1	2.2	2.0							1732.2	-
11	2.0	2.7	2.9	2.6						1411.3	-
12	2.1	2.0	2.3							1698.4	-
13	2.8	3.3	2.5	2.2	1.8	2.0				1335.7	1880.9
14	2.7	2.4	2.5	1.5	2.5	2.2	2.2			1591.7	1578.0
15	2.4	1.8	3.3	2.8	2.2	2.4				1385.9	1583.5
4 PM											
Cycle Number											
1	3.8									958.7	-
2	2.5	1.6	2.1	3.9	2.5	2.3				1430.0	1511.3
3	2.5									1464.6	-
4	2.5	3.7	2.2	2.4	2.5	2.9	2.9			1337.5	1303.7
5	3.9	1.8	2.2	2.4	2.2	2.2	1.7	2.6		1401.7	1640.5
6	1.7	2.0	2.6							1723.0	-
7	3.5									1032.7	-
8	4.7	2.4	2.3	2.1	1.4	1.5				1245.2	2480.2
9	4.8	2.0	2.6	1.8						1286.4	-
10	3.3	3.3	3.7	2.9	2.1	2.0	1.8			1100.5	1823.7
11	2.5	2.2	2.5	2.2	2.4	2.1	2.1	2.3		1525.1	1612.5
12	3.1	2.1	2.9	3.0	1.7	2.0	2.4			1296.5	1761.8
13	2.9	2.1	2.1	2.8	2.2	2.2				1453.2	1646.8

Table 15 presents a subset of headway and flow measurements. Data may not exist in every column of each row as there were a number of cycles that either did not

have a queue or the queue had less than nine vehicles. In addition, only 28 cycles were observed from the videos that were recorded during the original NGSIM experiment. The two sets of headway data (12 PM and 4PM) were statistically compared to determine the homogeneity between the two sets, and the appropriateness of combining the two datasets. The two-sample Wilcoxon Rank Sum (p= 0.791) and the Kolmogorov-Smirnov (p=0.881) tests both indicated that there was not sufficient evidence to reject the hypothesis that they represented the same distribution and thus for the following analyses the data were combined into a single dataset. In a final recommended implementation, the impact of calibrating models for each time periods separately could be explored, this will be discussed later in this chapter. The summary statistics for the combined startup and saturation flow measurements are presented in Table 16. Figure 46 presents a scatter plot of the field flow measures and their respective means (the red horizontal lines). The associated density plots pertaining to each set of flow measurements are presented in Figure 47

Table 16 Summary Statistics for Field Startup and Saturation Flow Measurements per Cycle

Statistic	Avg. Startup Flow (veh/hr/ln)	Avg. Saturation Flow (veh/hr/ln)
Min	959	1229
Mean	1408	1673
Max	1732	2480
Standard Deviation	182	281

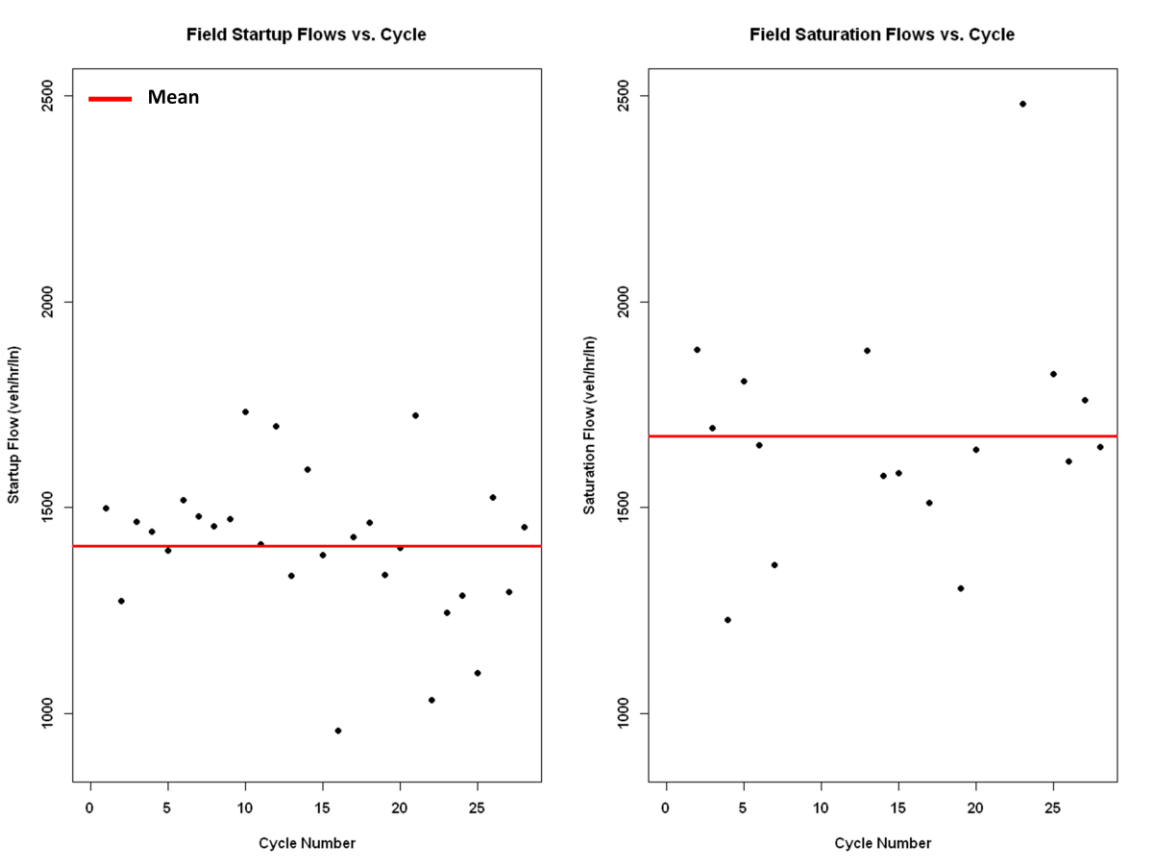


Figure 46 Field Startup and Saturation Flows versus Cycle Number

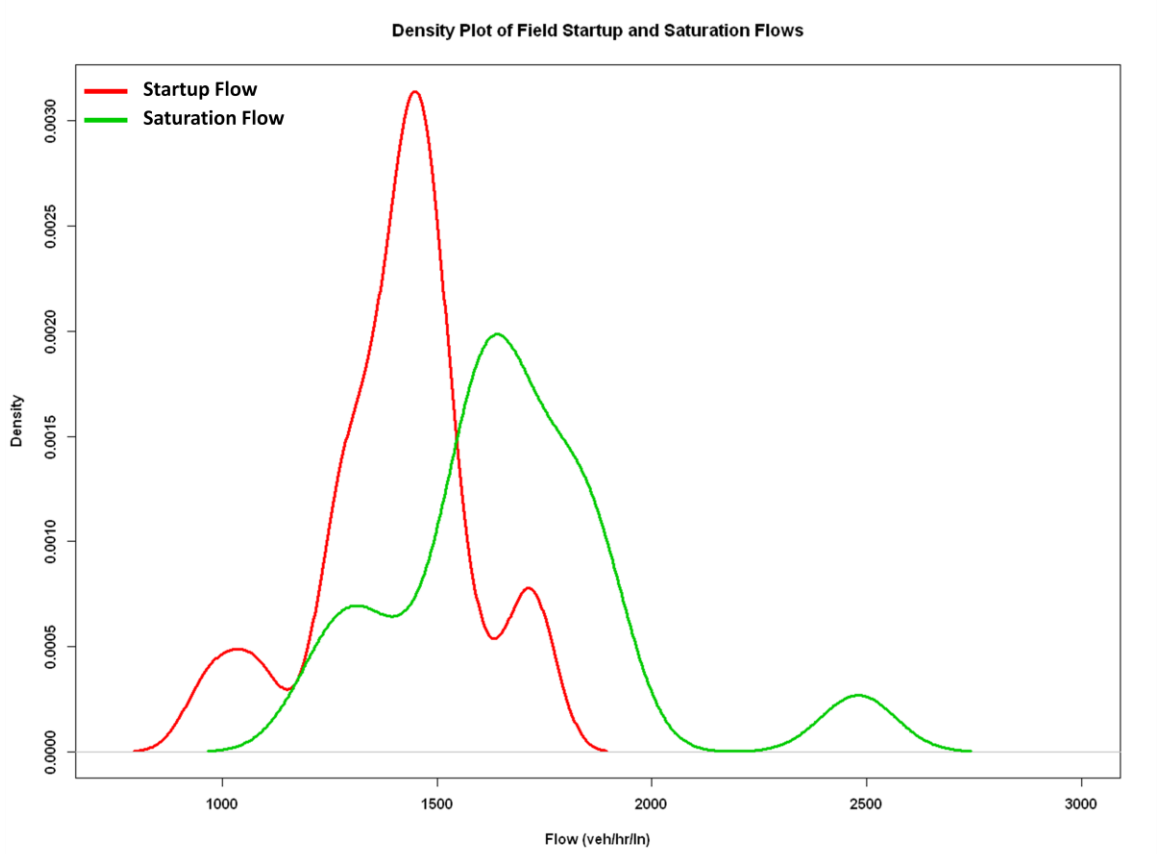


Figure 47 Density Plots of Field Startup and Saturation Flows per Cycle

As expected, it is seen that the startup headway distribution has a higher kurtosis and is shifted to lower values than the saturation headway distribution, representing the impact of start-up lost time. Similar headway and subsequent flow data were collected from the 1000 parameter set simulations of the study corridor. Table 17 presents summary statistics for data collected from all the parameter set simulations. Figure 48 presents the average startup and saturation flow estimates for each parameter set. Each point represents a simulated parameter set's average flow estimate and the horizontal line depicts the corresponding estimate from the field. This figure highlights that although there are parameter sets that produce estimates close to the field measurements, there are

a number of estimates that differ from field measures, by as much as 50%. Figure 49 and Figure 50 illustrate the density plots for startup and saturation flow per cycle from each parameter set simulation and from the field. This figure reinforces the large discrepancies between the flow estimates from the field and the simulated environment.

Table 17 Summary Statistics of Startup and Saturation Flow Measurements for Each Parameter Set

Statistic	Avg. Startup Flow (veh/hr/ln)	Avg. Saturation Flow (veh/hr/ln)
Min	960	798
Mean	1357	1433
Max	1886	2257
Standard Deviation	181	306

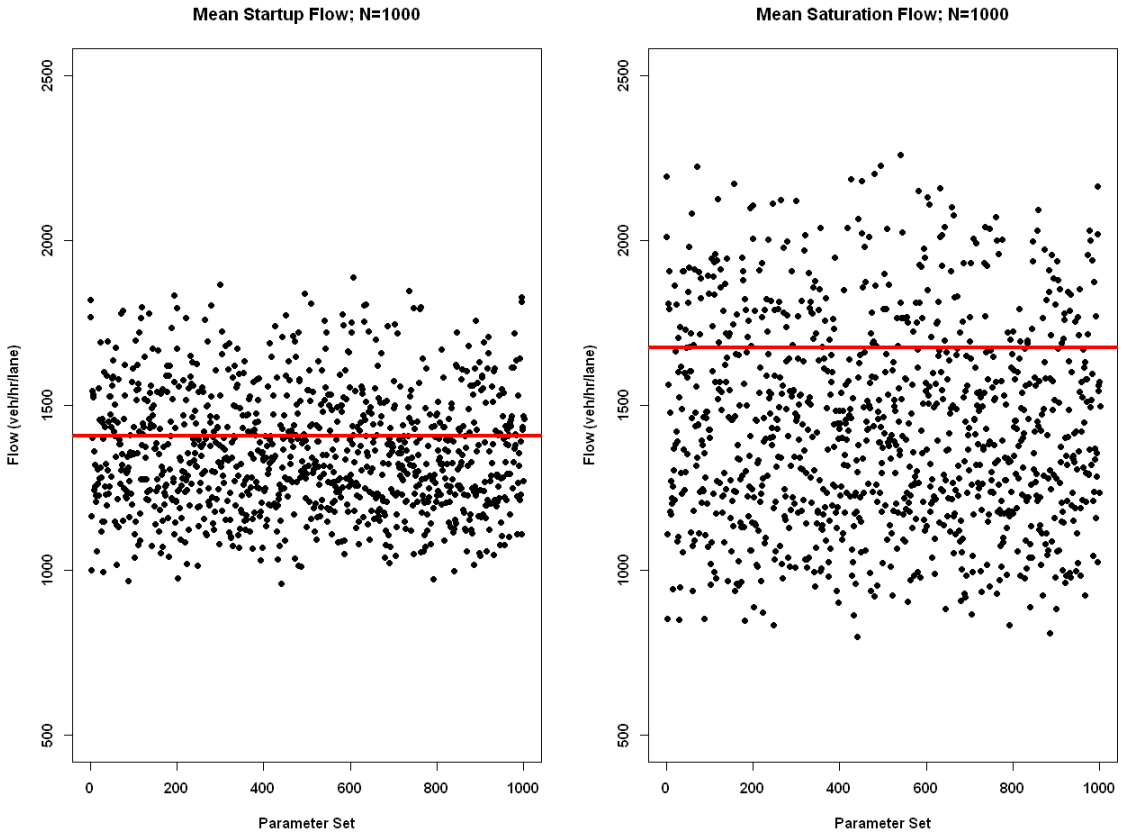


Figure 48 Startup and Saturation flow Estimate from Each Parameter Set Simulations

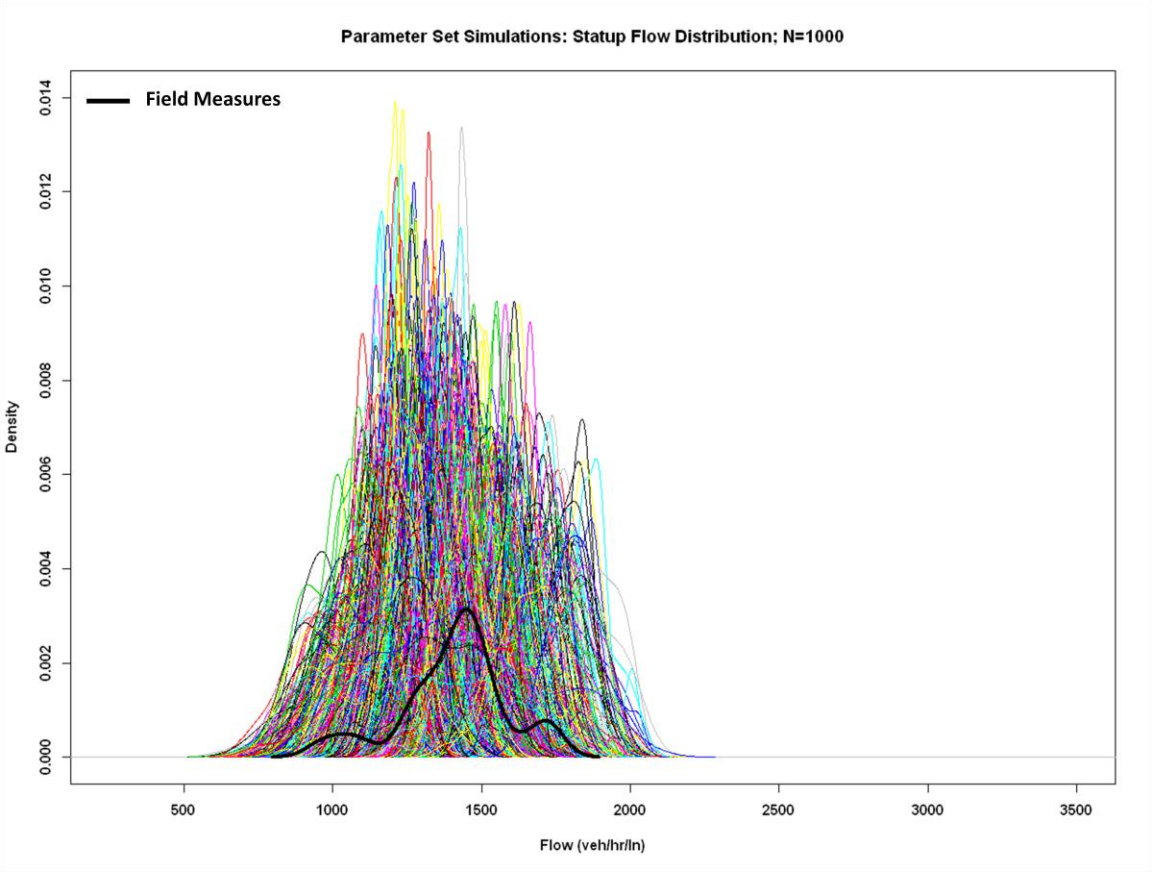


Figure 49 Density Plot of Startup Flow per Cycle

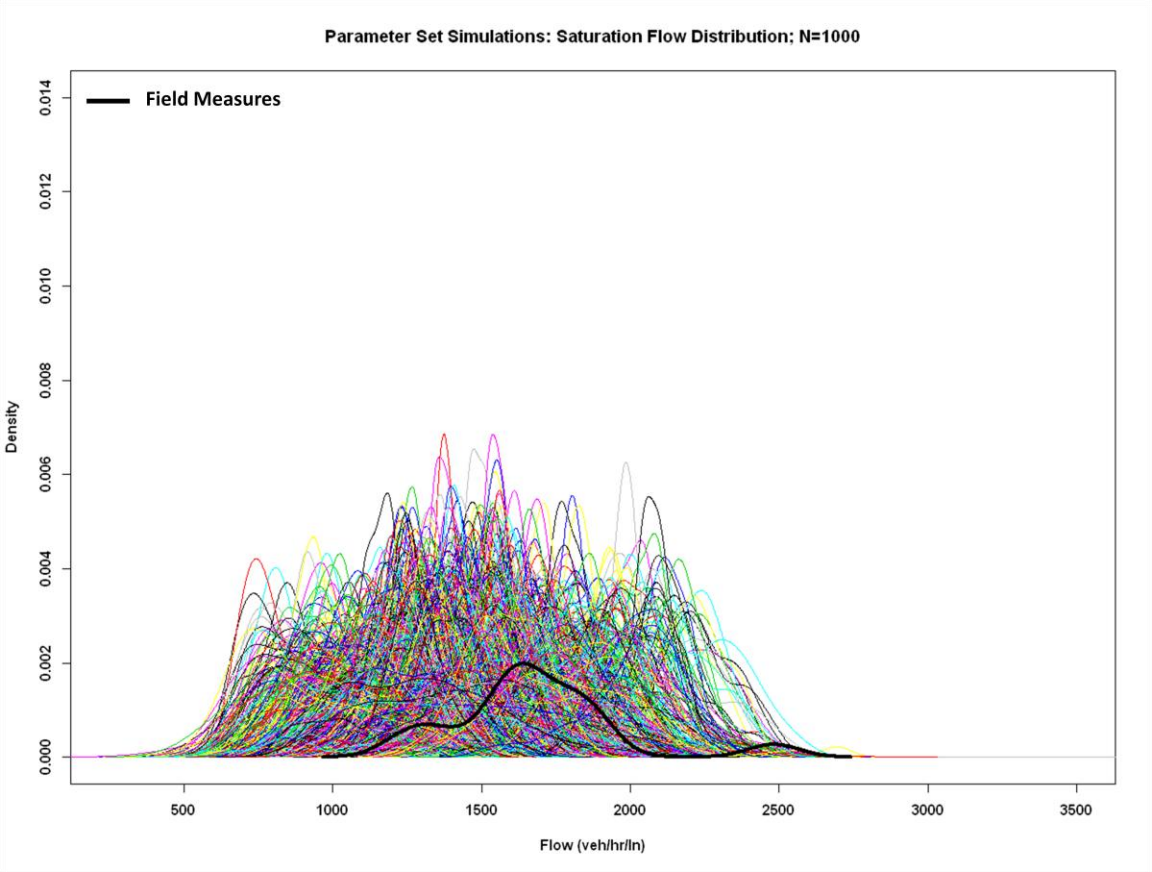


Figure 50 Density Plot of Saturation Flow per Cycle

To create the startup and saturation flow criterion, a reasonableness range was chosen to aid in the evaluation of whether or not a parameter set may potentially be considered calibrated. The reasonableness range was constructed by forming a 95% confidence interval around the mean flow values from the field. A challenge in constructing an appropriate confidence interval is the limited field data, particularly for the 6-9 vehicle group. In an attempt to address the negative consequences that may occur because of limited data, such as incorrect variance estimation, a bootstrap approach was taken to bolster the field data that will be used to inferences about field startup and saturation flow estimates.

Bootstrapping is a means of making statistical inferences in the presence of limited data. The bootstrap method involves the re-sampling of data, with replacement, in order to generate an empirical estimate of the entire sampling distribution of a statistic [54]. To carry out the bootstrap approach on the data in Table 15 10,000 re-samples were randomly drawn, with replacement, from the startup and saturation flow estimates. Appendix A presents five example resamples of both the startup and saturation flow estimates. The descriptive statistics from each of the 10,000 resamples will form the basis from which all inferences about field startup and saturation flow measurements will be made. Table 18 presents the summary statistics for the mean of the resamples as well as the average standard deviation for the respective resamples. With the bootstrap method addressing the dearth of field data, appropriate confidence intervals may now be developed to eliminate parameter sets.

Table 18 Summary Statistics of the Mean of Bootstrapped Field Flow Data

Statistic	Avg. Startup Flow (veh/hr/ln)	Avg. Saturation Flow (veh/hr/ln)
Min	1281	1409
Mean	1408	1674
Max	1546	1943
Standard Error	34	66
Avg. Standard Deviation	177	265

5.2.1.1.1 Forming Bootstrap Confidence Intervals

There are four common methods of forming confidence intervals when data are bootstrapped: normal approximation, percentile, bias corrected (BC), and percentile-*t* methods. Each of these methods were explored to produce an appropriate confidence interval. Table 19 presents the confidence intervals produced by each of these methods. A

summary of each method and how their confidence intervals were constructed is given in Appendix B. For a complete description of each method readers are encourage to consult Mooney and Duval (1993) [54]. Selecting a confidence interval method from the table below will be based on the width of the interval, its pros and cons, and the consistency of the intervals' end points.

Table 19 Bootstrap Confidence Intervals Method and the Intervals Produced

	Confidence Interval (CI)			
Method	Startup Flow		Saturation Flow	
	Lower Bound	Upper Bound	Lower Bound	Upper Bound
Normal Approximation	1341	1474	1544	1805
Percentile	1340	1473	1552	1811
Bias Corrected	1340	1473	1552	1811
Percentile *t*	**1342**	**1488**	**1499**	**1796**
CI From Raw Data	1337	1478	1529	1818

All four confidence intervals in Table 19 are similar. The interval produced by the normal approximation method was not chosen because its accuracy is dependent on the startup and saturation flow rates, per cycle, being normally distributed [54]. Although the field data in this case may not be rejected as normally distributed, according to Lilliefors normality test, requiring future data to be similarly distributed limits the applicability of the method being proposed and therefore reduces its desired level of robustness.

The confidence intervals produced by the percentile and the bias corrected method were also rejected. The confidence interval produced by the percentile method was eliminated for two reasons: 1) it performs poorly with small samples, and 2) it assumes that the bootstrap sampling distribution is an unbiased estimate of the statistic being evaluated [54]. Although neither of these reasons may limit the application of this

method to the current data, a robust approach to developing confidence intervals requires as few assumptions as possible. As for the bias corrected method, its interval was rejected because of it emphasis on a data point that may be unimportant to the data being studied, which becomes a greater issue when the data is skewed. [54]

The confidence intervals produced by the percentile-*t* method were selected as the reasonable range for simulated flow rates. The strength of this method is that it produces the most accurate estimates of a confidence interval from bootstrapped data. The disadvantage of this method is that it is computationally expensive because of the double bootstrapping that takes place. However, given the computing power that is available today, this disadvantage does not inhibit the usage of this method, especially for easily calculable statistics, such as means.

For a more complete summary of each method's advantages and disadvantages, see Table 20.

Table 20 A Comparison of Bootstrap Confidence Interval Methods [54]

Method	Advantages	Disadvantages
Normal Approximation	Similar to the familiar parametric approach; useful with a normally distributed estimator; requires the least computation (50-200 re-sampling events)	Fails to use the entire bootstrap sampling distribution; requires parametric assumption about distribution of population characteristic
Percentile	Uses the entire bootstrap sampling distribution; allow population distribution to be asymmetrical; invariant to transformation	small samples may result in low accuracy; assumes bootstrap sampling distribution is unbiased
Bias Corrected	All of those of the percentile method; allow for bias in the bootstrap sampling distribution; *z*-value can be calculated easily from bootstrap distribution	Requires a limited parametric assumption
Percentile-*t*	Highly accurate confidence intervals in many cases; handles skewed population distribution better than the percentile method	Not invariant to transformation; computationally expensive with double bootstrap

As seen from Table 20, the percentile-*t* is not disadvantaged by any parametric assumptions. This enables the method to be used to produce acceptable intervals for datasets with a wide range of characteristics. In comparing the current data set to the selected intervals, Table 19, a window that is 10% and 18% of the field's mean startup and saturation flow estimates, respectively, are reasonable flow ranges for the study corridor. This further underscores that the selected range is a feasible range to eliminate parameter sets from being considered as calibrated models with unreasonable startup and saturation flow estimates.

5.2.1.1.2 Applying the Startup and Saturation Flow Criteria

Applying the saturation flow criteria is a two-phase process. The first phase selects parameter sets that produce flow estimates that are within the selected confidence intervals for each of the vehicle groups. The second phase identifies parameter sets that produce saturation flow estimates that are within both sets of confidence intervals, and this becomes the new list of candidate parameter sets. From the first phase, 253 parameter set simulations produced startup flow estimates within the respective confidence interval while 255 were within the confidence interval of the saturation flow estimates. The second phase resulted in 159 parameter sets that produced flow estimates within both set of confidence intervals. Figure 51 – Figure 53 illustrates these two phases. The green horizontal lines in each of these figures represent the percentile-*t* 95% confidence interval for the respective flow measures, while the red line represents the respective mean flow value. Figure 51 shows all the parameter set simulations' estimates of startup and saturation flow. Figure 52 illustrates the flow measures that are within the respective intervals. Finally, Figure 53 presents the final set of parameter set simulations that produced flow values that were within both sets of confidence intervals.

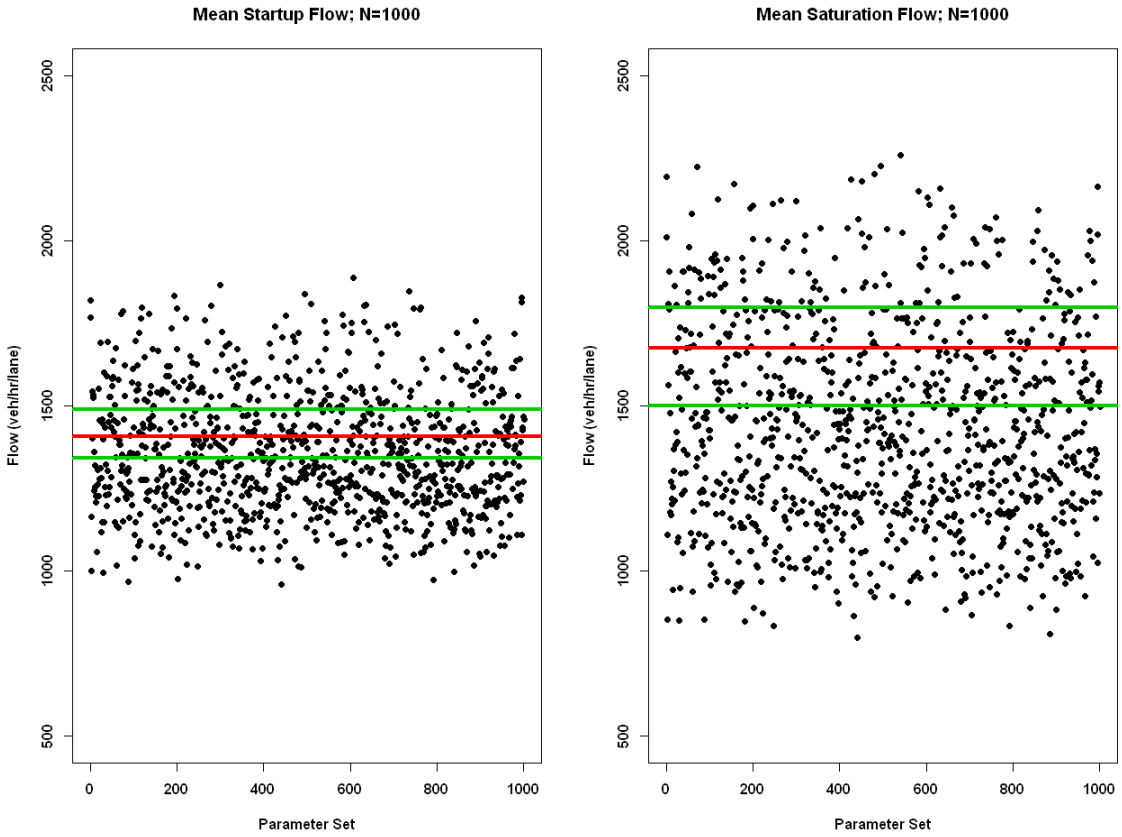


Figure 51 Average Field and Simulated Startup and Saturation Flows with the 95% Confidence Intervals

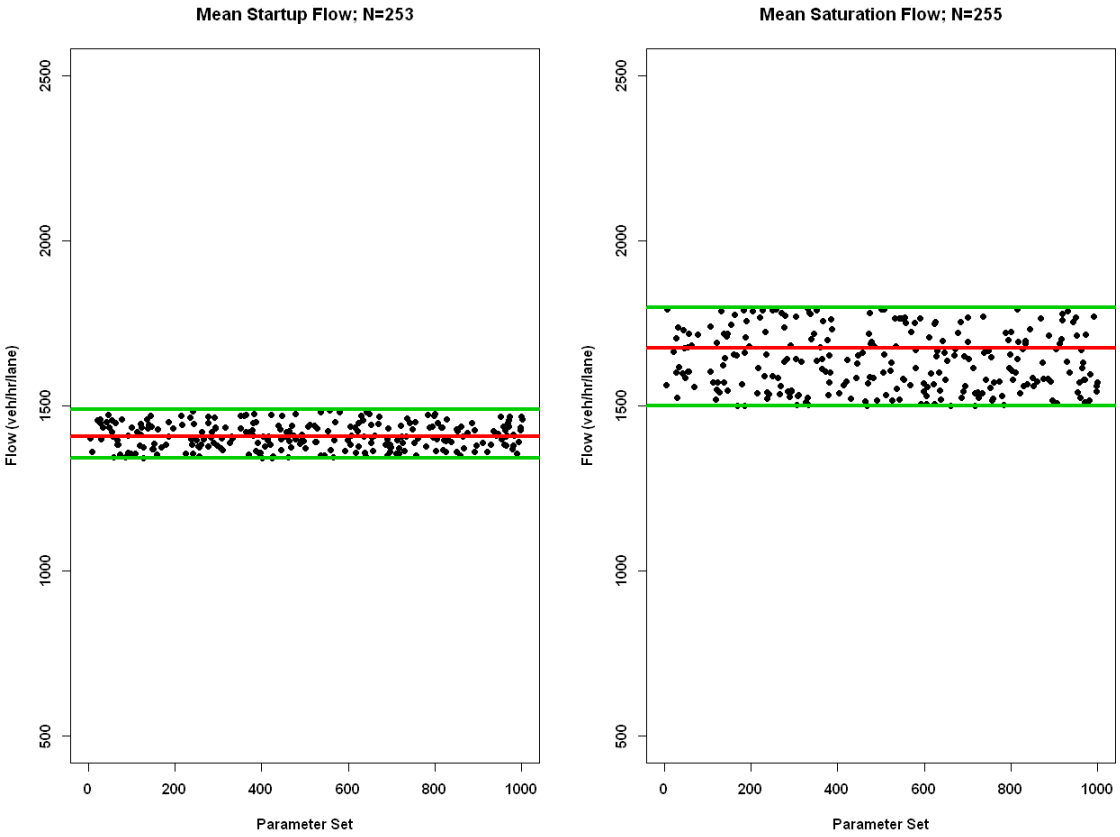


Figure 52 Average Field Flow Estimates and Simulated Flow Estimates within the 95% Confidence Intervals

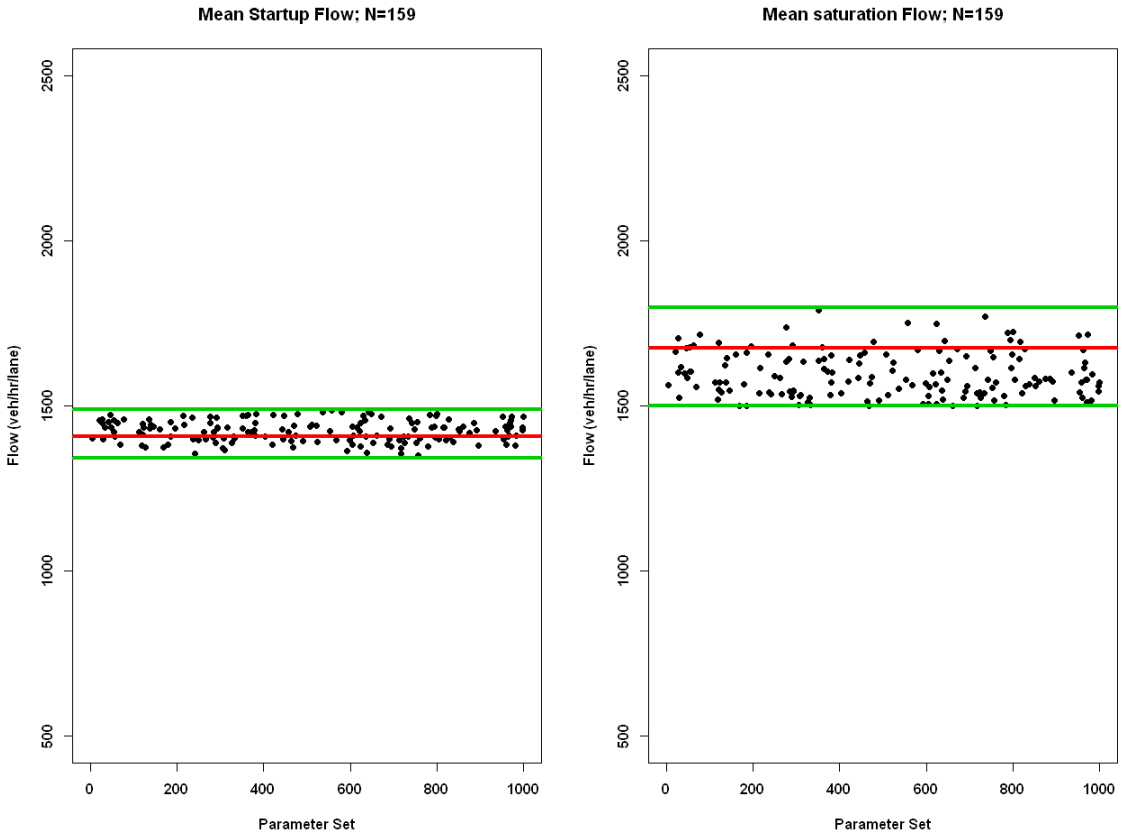


Figure 53 Final Set Flows that Satisfy Saturation Flow Criteria

Upon applying the final step of the flow criteria, Figure 53 presents a rather significant result when comparing the startup and saturation flows to their respective means. One hundred estimates of startup flow were *greater* than or equal to the mean field startup flow of 1408 veh/hr/ln. While, 137 saturation flow estimates that are *less* than the mean field saturation flow of 1673 veh/hr/ln. The working hypothesis for this disproportionate selection of startup and saturation flow estimates, relative to their means, is that there appears to an inverse relationship between the final two sets of selected flow measurements. In addition, it is also possible that this inverse relationship

attempts to maintain a ratio of saturation to startup flow, which is influenced by the same ratio in the field.

Results from preliminary investigations into the theorized inverse relationship and ratio have not led to a rejection of the working hypothesis. In terms of the inverse relationship between the two sets of flow measurements, the calibration parameters that are most influential to these measures were examined. The average standstill distance parameter (*ax*) is the drives the startup flow estimates while the safety distance parameter (*bx*, which is a function of *bx_add* and *bx_mult*) drives saturation flow estimates. Comparing values of these parameters, for a selected parameter set, does highlight this inverse relationship. Additionally, this inverse relationship also tends to bring about a convergence on a small range of values for *d* – a measure of average distance between vehicles.

The inverse relationship and the converging values of *d* seem to be parallel with the inverse relationship between startup and saturation flow, and the convergence on a particular ratio of the two. Figure 53 illustrates the inverse relationship (relative to their mean values for the field). The ratios of saturation to startup flows from each of the selected parameter set were calculated. For this set of ratios, the minimum, mean and maximum values were 1.02, 1.12, and 1.21 respectively. The mean ratio from the field is 1.19. The comparison of these ratios seems to suggest that selected parameters sets are attempting to converge on the field value of the ratio (within the given range of acceptable flows). This was further highlighted as a possibility as the minimum, mean and maximum ratios from the all (1000) candidate parameter set simulations were 0.76, 1.05, and 1.30 respectively.

Future efforts will continue the investigation into this working hypothesis. In particular, these efforts will be spent developing a comprehensive understanding of the suggested inverse relationships and the convergence to a single value or small range of values. Section 5.3, which presents a few observations when trying to select parameter values for model calibration, will lend some additional insight into this investigation. Additionally, the physical interpretation of the acceptance or rejection of the hypothesis will be examined.

To further demonstrate the reasonableness of the selected confidence intervals, the density plots of startup and saturation flow from the final selection of parameter sets are presented in Figure 54 and Figure 55. This figure indicates that the selected parameter sets not only reflect the mean start-up and saturation flows from the field but also highlight the elimination of those that are the poorer fits according to the distribution of flow per cycle.

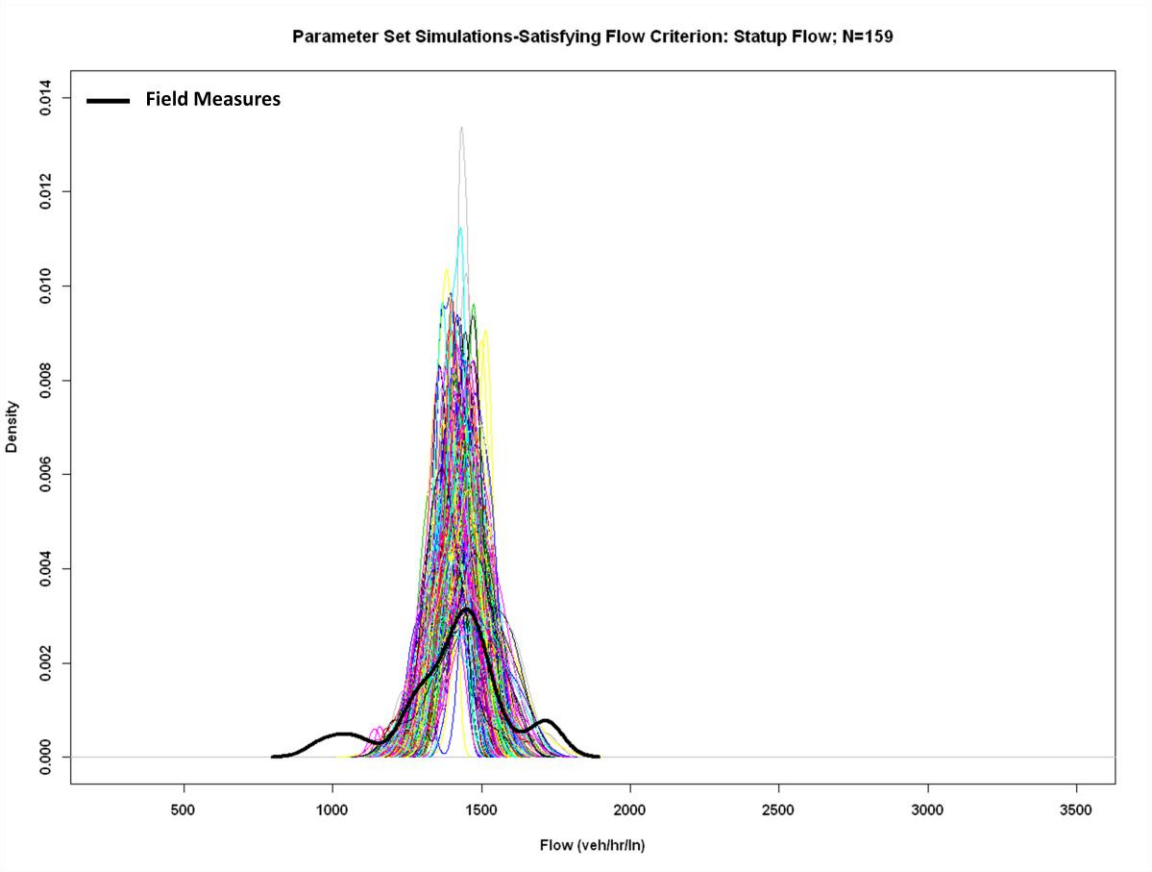


Figure 54 Density Plot of Parameter Set Simulations that Fit Flow Criteria – Startup Flow

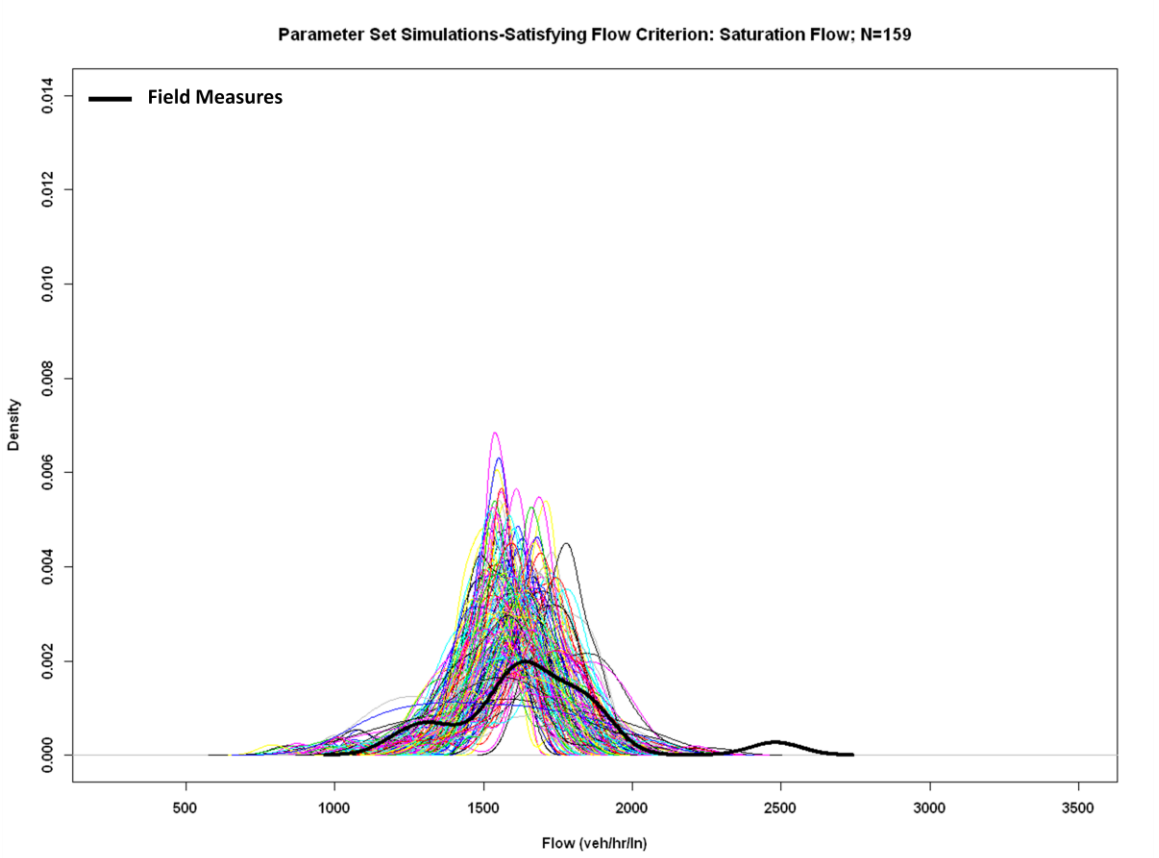


Figure 55 Density Plot of Parameter Set Simulations that Fit Flow Criteria – Saturation Flow

These remaining 159 parameter sets are now the complete set of candidates from which final calibrated models will be selected. To identify calibrated models from this set of remaining parameter sets, part two of the criteria process, the statistical evaluations of distributions of travel time, will be applied. The following sections will detail the development of how the performance measures from each parameter set simulation will be evaluated statistically to select calibrated parameter sets.

5.2.1.2 Statistical Evaluation Criteria

For statistical evaluation criteria travel time distributions from each parameter set simulations were examined. Parametric and non-parametric statistical tools were explored to compare field and simulated travel time estimates. Parametric tools are more commonly used for these types of comparisons. However, the limitation of parametric tool is that they require data to fit a known distribution, typically a normal distribution, and field data often do not fit the appropriate distributions. Therefore, to ensure proper comparisons of field and simulated data, non-parametric tools were used.

Two sets of non-parametric tools were used to establish this statistical calibration criteria, one is a general distribution comparison to primarily determine the homogeneity between field and simulated travel estimates. The other set of tools performed a more stringent comparison of travel time distributions. With these tools working in tandem, parameter set simulations that produce accurate estimates of field travel distributions will be considered as calibrated replicates.

5.2.1.2.1 General Distribution Comparison

A number of non-parametric tests were examined to evaluate the general differences in the distributions of field and simulated travel times. The Wilcoxon rank sum and Mann-Whitney tests were identified as appropriate tools for such an evaluation. When comparing populations, the rank-sum test has as its null hypothesis that the "two population are equal – $F_x(x) = F_y(x)$ [55] p. 127. In other words, the rank sum tests tries to determine "whether the groups are homogeneous or one group is 'better' than the other" [55] p. 127. To perform this test, ranks are assigned to each data point in each sample. These ranks were determined by the data point's position after both samples are

combined and ordered. The summation of these ranks, for each sample, will then indicate whether both samples have the same distribution. If the sum of ranks of both samples is the same, then this indicates that both populations have the same distribution. More formally, the rank sum test statistic is given below in Equation 2.

$$\boldsymbol{W = \sum_{i=1}^{n} iS_i(X,Y)} \qquad \textbf{(2)}$$

where $S_i(X,Y)$ is an indicator function which is equal to 1 when the i^{th} ranked observation is from the first sample and 0 otherwise; and $n = n_1 + n_2$ [55]

The Mann-Whitney test was also chosen to evaluate the general differences in the distributions of field and simulated travel times. The Mann-Whitney test was selected as its function is "to find differences in two populations" [55] p. 129. Given the similarity in intentions between this test and the rank sum test, and the equivalency of the Mann-Whitney test statistic (U – Equation 3) and that of the rank sum test; the Wilcoxon-Mann-Whitney (WMW) test was chosen to evaluate the general differences between field and simulated travel time distributions. The equivalency of the Wilcoxon test statistic and that of the Mann-Whitney test will not be presented here. However, for interested readers, please consult Kvam and Vidakovic (2007) [55].

$$\boldsymbol{U = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} D_{ij}} \qquad \textbf{(3)}$$

where n_1 and n_2 is the number of elements in the first and second sets respectively and D_{ij} is the difference between ranks.[55]

The Wilcoxon-Mann-Whitney test was implemented using R, a language and environment for statistical computing and graphics [56]. More specifically, the WMW test was applied using the *wilcox.exact()* command, which is a member of R's "exactRankTests" package [57]. One of the outputs from the *wilcox.exact()* command is a

p-value that is used in the decision to accept or reject the null hypothesis, H_o, which is in essence states that the field and the simulated travel time distributions are likely equivalent. A calibration criterion based on the WMW *p-value* is the rejection of H_o, and subsequently a model parameter set, when the *p-value* is ≤ 0.01. This criterion is then paired with another that is based on a more stringent comparison travel time distribution.

The need for a more stringent comparison of travel time distributions is due to one of its fundamental principles of the WMW test. Because the WMW relies on ranks, the evaluation of travel time distributions is based on relative differences amongst the data points versus the magnitude of those differences. By not basing the comparison on the magnitudes of these differences, it weakens the distribution comparison between field and simulated measures. At times, the WMW test is used to compare differences in medians of two distributions. However when conducting such a comparison, it is assumed that these distributions are identically shaped. Irrespective of the assumption that is made, the execution of the WMW test does not change, only the interpretation of the results.

The ability for the WMW to compare medians will play a role in the formulation of this criterion. In particular, in the final stage of this criterion, the results from the WMW test will be interpreted with respect to its capability to compare medians. This is because of combining the WMW test with two distribution tests that outputs similarly shaped distributions. This combination therefore addresses the similarly shaped distribution assumption and therefore allows the WMW test to evaluate medians appropriately. The following section will present the details of the two distribution tests

that evaluates distribution shapes. These two test will bolster the distribution comparisons and subsequent statistical evaluation criteria.

5.2.1.2.2 Secondary Distribution Comparison

Similar to selecting a test to compare the general form of the distributions of travel time measurements, a number of additional tests were also examined to conduct a more stringent comparison of these distributions. Given the multimodal travel time distributions, with hard to discern parametric descriptors, the Kolmogorov-Smirnov (KS) non-parametric test was selected to compare the distribution of travel time datasets. The KS test was considered as a secondary distribution test as its test statistic "is the basis of many nonparametric goodness-of-fit for distributions" [55] p. 78. In addition, unlike the WMW the KS test does take into account the magnitude of the differences between the data points of the samples being compared. However, given the data at hand, the KS test was applied in conjunction a less formal, non-parametric heuristic form fit (HFF) test to select replicate runs that produced travel time distributions that were *more* similar to those obtained from the field.

The HFF test was included to provide an alternative distribution comparison method whose assumption(s) were not violated by the data at hand, unlike the KS test. One of the assumptions of the KS test is that data being analyzed is continuous [48] [55]. The field travel time is not continuous (see Figure 56 - Figure 59) and it is expected that there will be circumstances under which replicates do not produce continuously distributed data. The lack of continuous travel times is in part due to signalized intersections not affording the realization of some travel time values. Although the KS test can still be applied to discontinuous data, inferring from possibly incorrect results

could prove costly. Incorporating the HFF test is an attempt to mitigate potential incorrect inferences and subsequent elimination of possible calibrated replicates. It is in this vein that both the KS and the HFF test will be used to evaluate distribution fits.

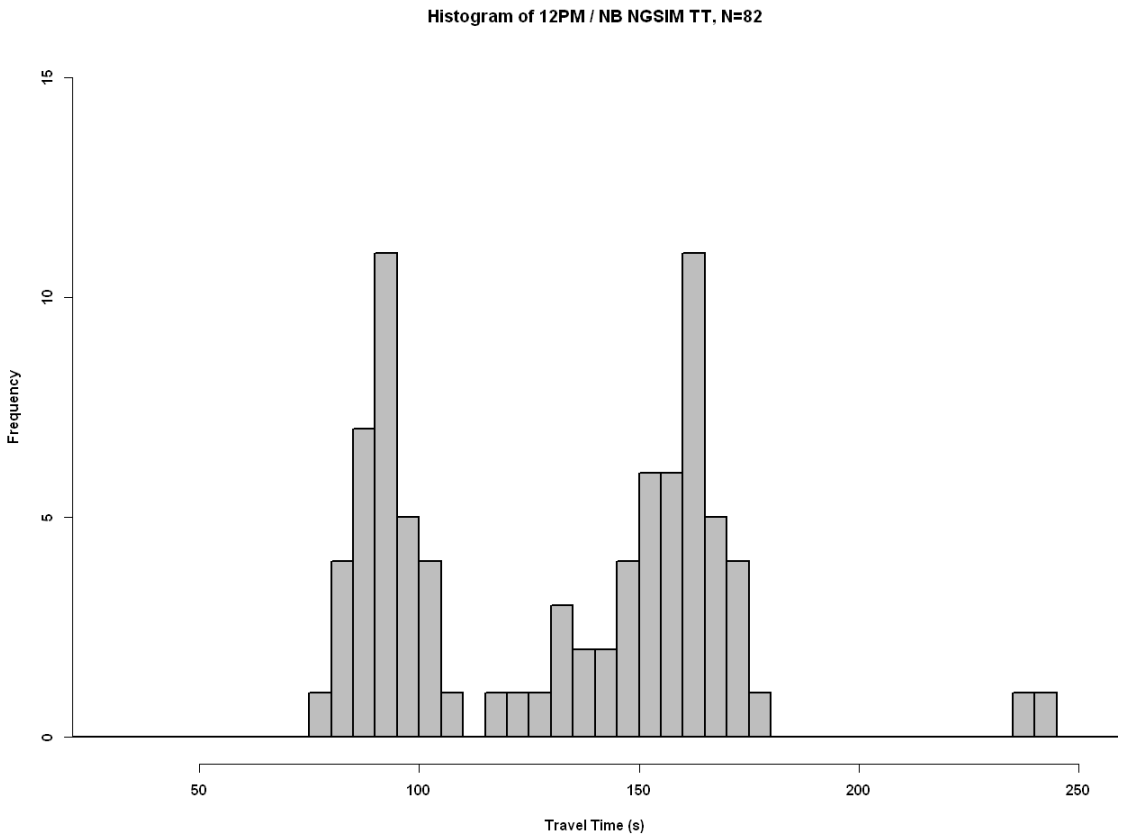
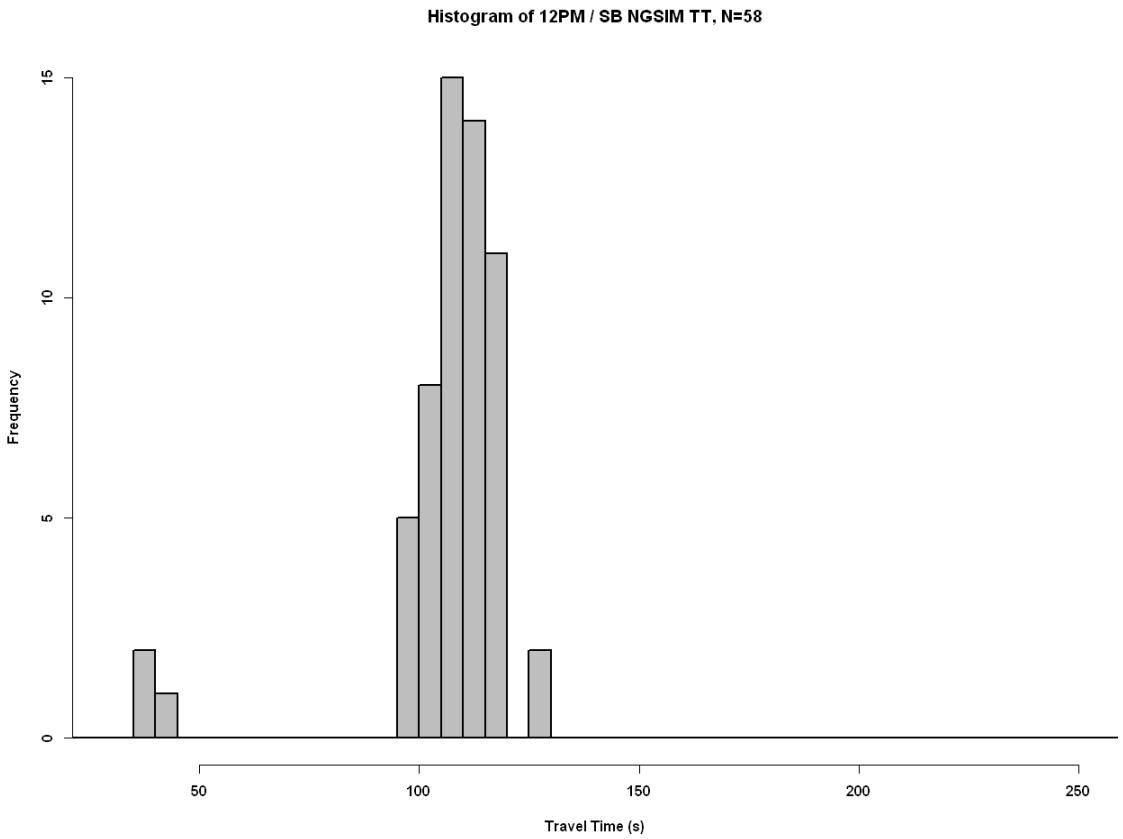


Figure 56 Demonstration of Discontinuous NGSIM Noon TT – Northbound



(b)

Figure 57 Demonstration of Discontinuous NGSIM Noon TT – Southbound

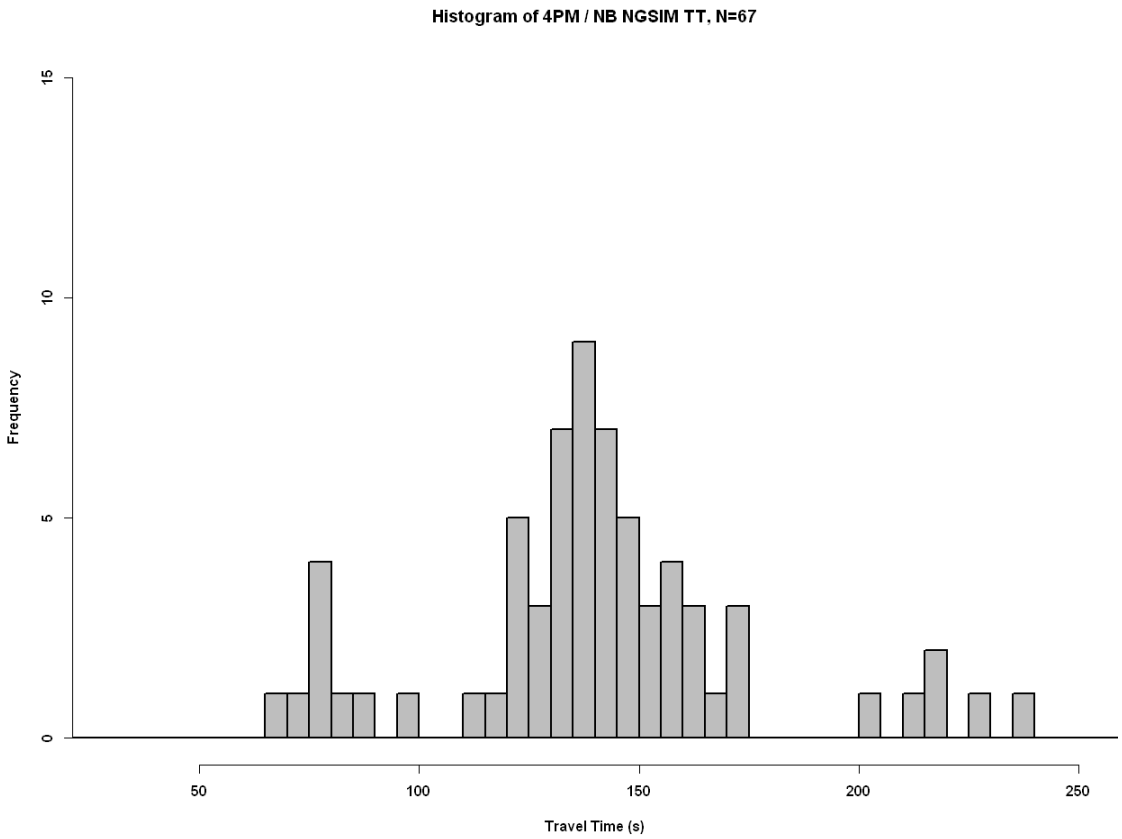


Figure 58 Demonstration of Discontinuous NGSIM Evening TT – Northbound

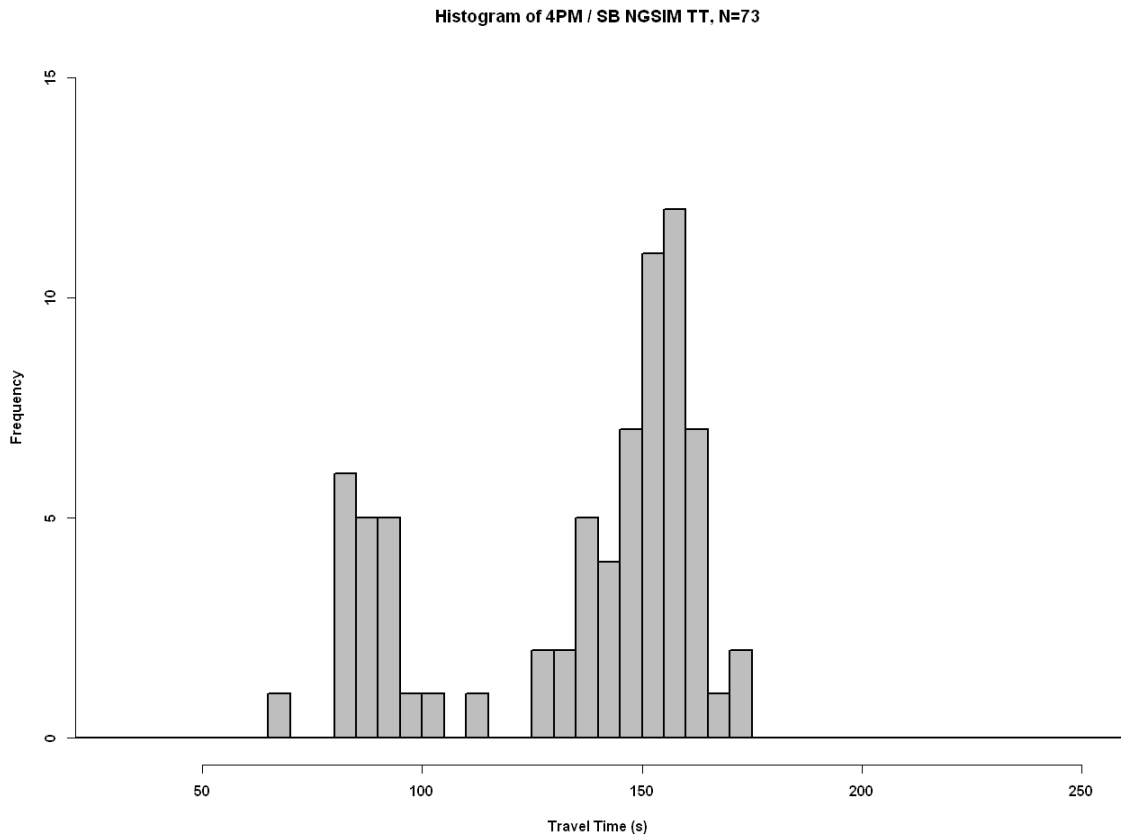


Figure 59 Demonstration of Discontinuous NGSIM Evening TT – Southbound

The formulation of the KS test statistic, D_n, explains the need for continuously distributed data. D_n is that the largest vertical distance between two cumulative distribution functions (CDF), say, $F_n(x)$ and $\dot{F}_n(x)$, for all values of *x*. In the context of this research, $F_n(x)$ and $\dot{F}_n(x)$, represent the CDF of the field and simulated data respectively. Formally

$$\boldsymbol{D_n = sup\{|F_n(x) - \dot{F}_n(x)|\}} \qquad \textbf{(4)}$$

where *sup* yields the smallest value from a set of values that is greater than or equal to all value in the set. [48]

A graphical representation of D_n is presented in Figure 60. This figure compares two sample CDFs and "*D*" is the largest vertical distance between the two. Applying the meaning of D_n to the selection of a calibrated versus uncalibrated model, the larger the magnitude of D_n for a given comparison of model versus field data, the less likely that that model will be considered as a calibrated model. To formalize the calibration criterion base on the KS test, the model comparison whose D_n value corresponds to a *p-value* of ≤ 0.01 will result in a rejection of H_o. The null hypothesis in this case states that there is insufficient evidence to suggest that a parameter set's simulated distribution of travel time estimates is different from same distribution obtained from the field. The rejection H_o for a particular travel time segment removes that parameter set from being considered as a possible calibrate model for a particular period and travel direction. The KS test was implemented in the R environment using the *ks.test()* command that includes an associated *p-value* as a part of its output [56].

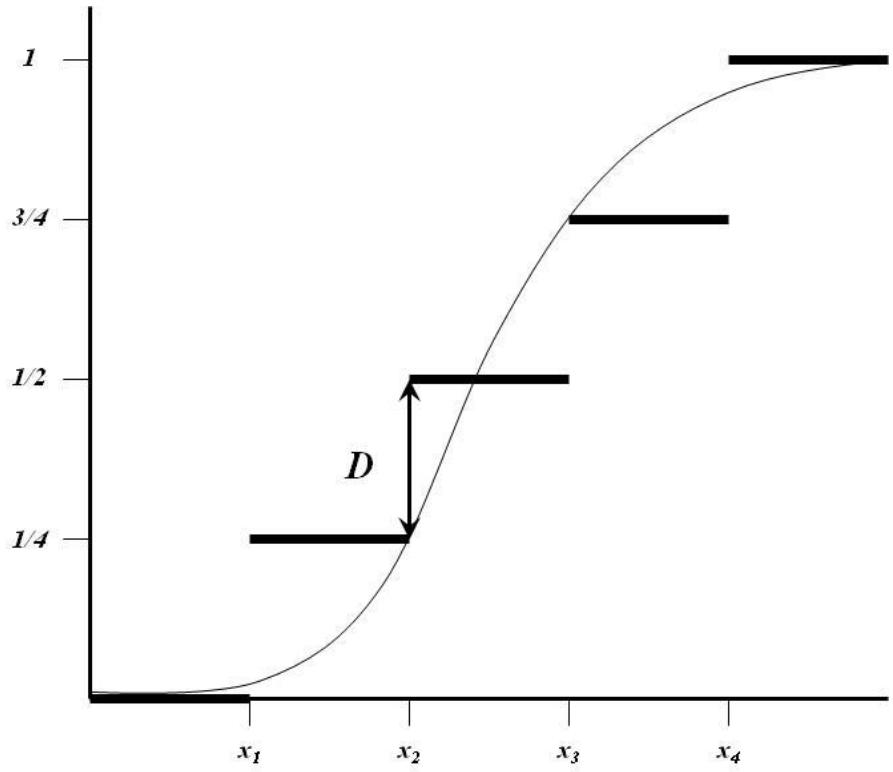


Figure 60 The meaning of the Kolmogorov-Smirnov Test Statistic, D_n [48]

To limit erroneous inferences from the KS test, particularly in the instances when the data being compared is discontinuous, the HFF test is devised to compare the rate of change of the CDF of the two distributions. A test statistic, *H*, is created and is defined as the sum of squares of the difference between the rate of change between the CDF of the field and simulated data, $F_n(x)$ and $\dot{F}_n(x)$, respectively. Mathematically,

$$\boldsymbol{H = \sum \left(\frac{dF_n(x)}{dx} - \frac{d\dot{F}_n(x)}{dx} \right)^2} \qquad \textbf{(5)}$$

There are no assumptions associated with the HFF test. To make inferences from the HFF test consider the magnitude of *H*. The smaller the magnitude of *H*, the more likely that the parameter set simulation being compared will be calibrated model. A disadvantage of this method is that it does not take into account shifts along the x-axis,

i.e. differences in central tendencies (mean/median). This is concerning as a for a given comparison, *H* having a value equal to or close to zero does not necessarily mean that the results from a parameter set fits field data. The only definitive statement that may be made is that the shapes, or forms, of the two distributions, including its modal characteristics, are similar. See Figure 61 which presents two similarly shaped CDFs, where $H \approx 0$, with very different means. To ward against selecting a replicate that only fits the shape of the distribution of the field data, the HFF test is paired with the MWW test to compare the mean of the two distributions.

The heuristic nature of the HFF test also aids in selecting a calibrated model. For a given travel time segment 1000 values of *H* will be calculated. Since there is currently no *p*-value for this test statistic, parameter set simulations that produce *H* values in the bottom half of the range of H-values are considered as calibration model candidates.

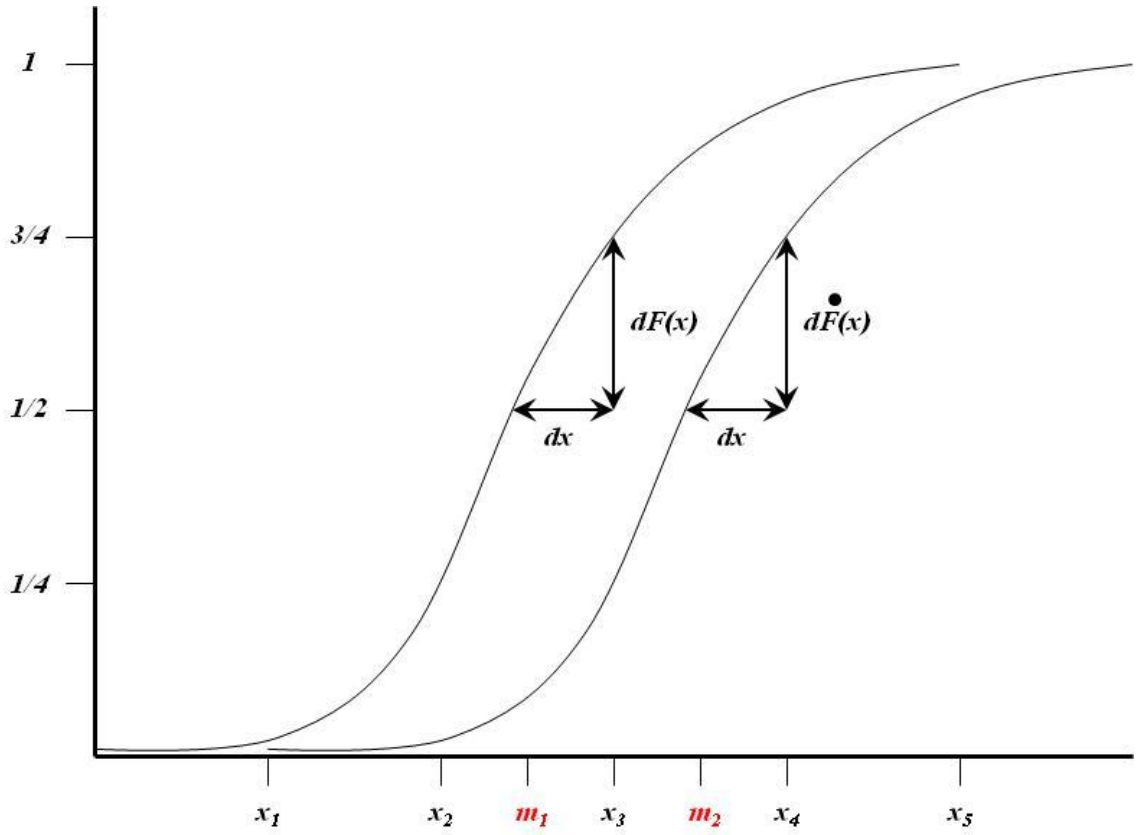


Figure 61 Similarly Shaped CDFs (equivalent *H*s) but Different Means ($m_1 \neq m_2$)

The HFF test was also implemented using R. In comparing two distributions, the H value is calculated by first obtaining the CDF for each and delineating 1% (or 0.01 probability) intervals. Each percentage change in the CDF corresponds to a change in travel time. For two distributions, the difference between the respective changes in travel time, per percent change, is squared and summed. This is the H value for the comparison of two distributions. Below is the pseudo-code that was used to populate an array, *H*, with all the values of the HFF test statistic for the comparison of field versus simulated travel time measures, for each of the 1000 parameter sets.

Pseudo-code:

```
for (i = 1 to number of replicate)  #number of replicates = 1000 in this case
  {
    for (j to 100 %)
      {
        squared error[j] =(field_tt_cdf[j+1]-field_tt_cdf[j])- (replicate _i_tt_cdf[j+1]-
        replicate_i_tt_cdf[j])^2
      }
    H[i]=sum(squared error)
  }
```

The statistical evaluation criteria uses the WMW, KS and HFF tests to thoroughly examine the field and simulated travel time distributions. All three statistical tests were used to determine which parameter set satisfy the statistical evaluation criteria. These tests were only performed on the parameter set simulations that satisfied the saturation flow criteria from above. The results from these tests will further determine which parameter sets are calibrated. The following steps outline the application of these tests and how parameter sets that satisfy the statistical evaluation criteria are selected:

1. Conduct the WMW test to perform a general comparison of field and simulated travel time distributions.
2. Retain parameter set simulations whose WMW test yielded *p-values* ≥ 0.01. This will be a set of parameter sets denoted by $M_n|U$.
3. Conduct the KS test to compare the distributions of field and simulated travel times.
4. Retain parameter set simulations whose KS test yielded *p-values* ≥ 0.01. This will be a set of parameter sets denoted by $M_n|D$.
5. Conduct the HFF test to compare the shapes of the distributions of field and simulated travel times.

6. Retain parameter set simulations whose *H* value is in the bottom half of the range of *H*-value. This will be a set of parameter sets denoted by $M_n|H$.
7. To obtain the set of parameter set simulations that satisfy the statistical evaluation criteria, carry out the following set operation

$$M_f = M_n|U \cap (M_n|D \cup M_n|H) \qquad \textbf{(6)}$$

The parameter set simulations that satisfy both the KS and HFF tests were combined as they evaluate the same characteristic of the data – the shape of its distribution. It will be seen that for described experiment that most models belonging to $M_n|H$ were already a part of $M_n|D$ except for a few instances. The union of these two sets facilitated the inclusion of simulated data that may have been otherwise excluded. The exclusion of such datasets may have been the result of a violation of a KS test assumption and/or the inability to provide a large enough *p*-value to be included in $M_n|D$, despite having similarly shaped distributions. In other words, $M_n|H$ and $M_n|D$ were combined to further minimize the probability of committing a Type I error, regarding distribution shapes.

The set operation in the seventh step identified parameter set simulations that produced travel time distributions that have similar shapes and medians as those from the field. The union operation identified parameter sets that provide similarly shaped travel time distributions. When performing the WMW test on similarly shaped distributions, their medians are being compared, as alluded to earlier. Therefore, the intersection operation identified parameter set simulations that produced travel time distribution with similar medians. The above seven steps bolstered part two of the two-phase, facilitating the appropriate selection calibrated parameter sets.

These seven steps were carried out for each time period and direction of travel in the NGSIM data set (12PM and 4PM, Northbound and Southbound). Figure 62 - Figure 68 illustrate the resulting sets throughout the process of selecting a calibrated parameter set. For illustrative purposes, only the resulting density from the noon-northbound travel time segment is presented in Figure 62 - Figure 64. For the complete set of resulting density plots please see Appendix C.

The first set of parameter sets simulations were eliminated after applying the WMW test. The remaining parameter set simulations had sufficiently similar distributions that there was insufficient evidence to reject H_0. Figure 62 presents travel time density plots for the models that belong to the set $M_n|U$. Similarly, the KS and HFF tests were conducted and Figure 63 and Figure 64 present the density plots for parameter set simulations belonging to $M_n|D$ and $M_n|H$ respectively. The final set of parameter set simulations that were obtained after applying Equation **(6)** to the above sets. These parameter set simulations, Figure 65 - Figure 68, are deemed calibrated – according to the two-part criteria process. Table 21 presents a summary of the number of parameter set simulations that were eliminated as the various calibration criterion were applied.

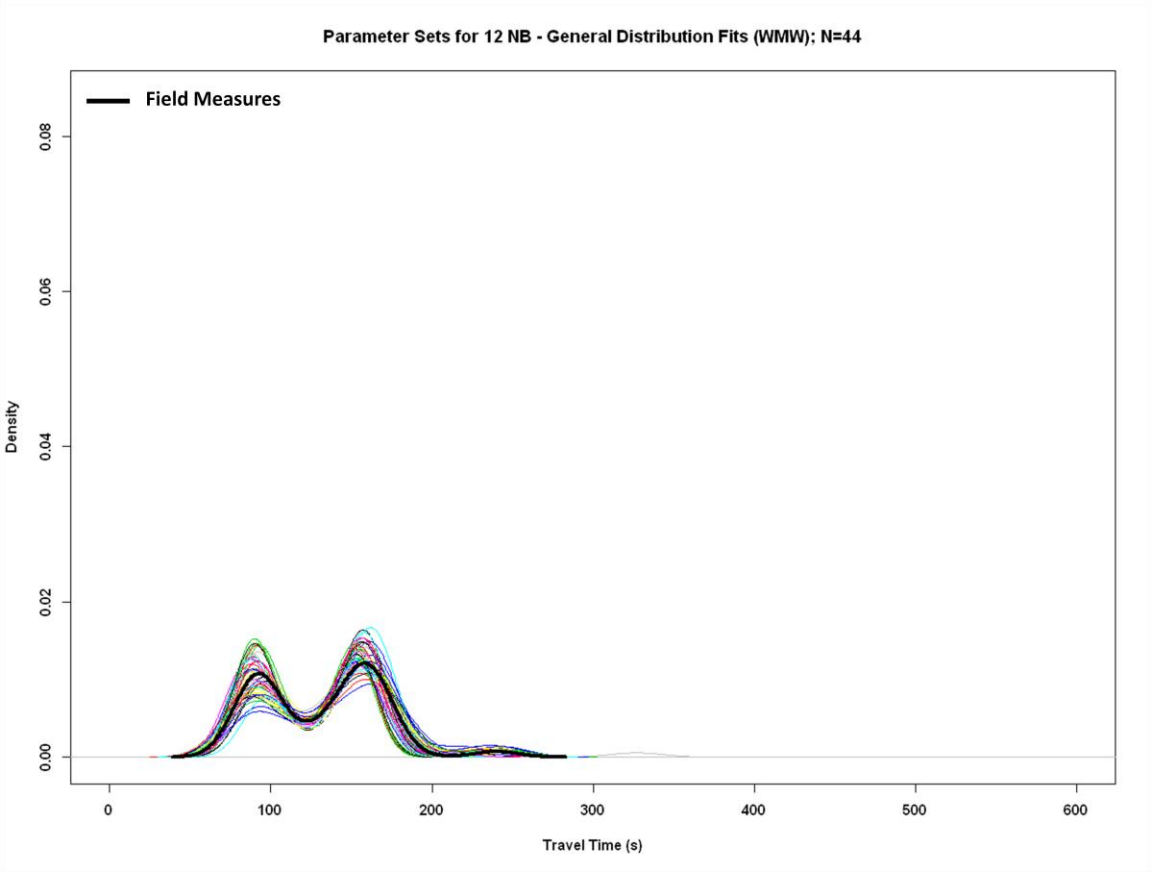


Figure 62 Models from WMW Test, $M_n|U$

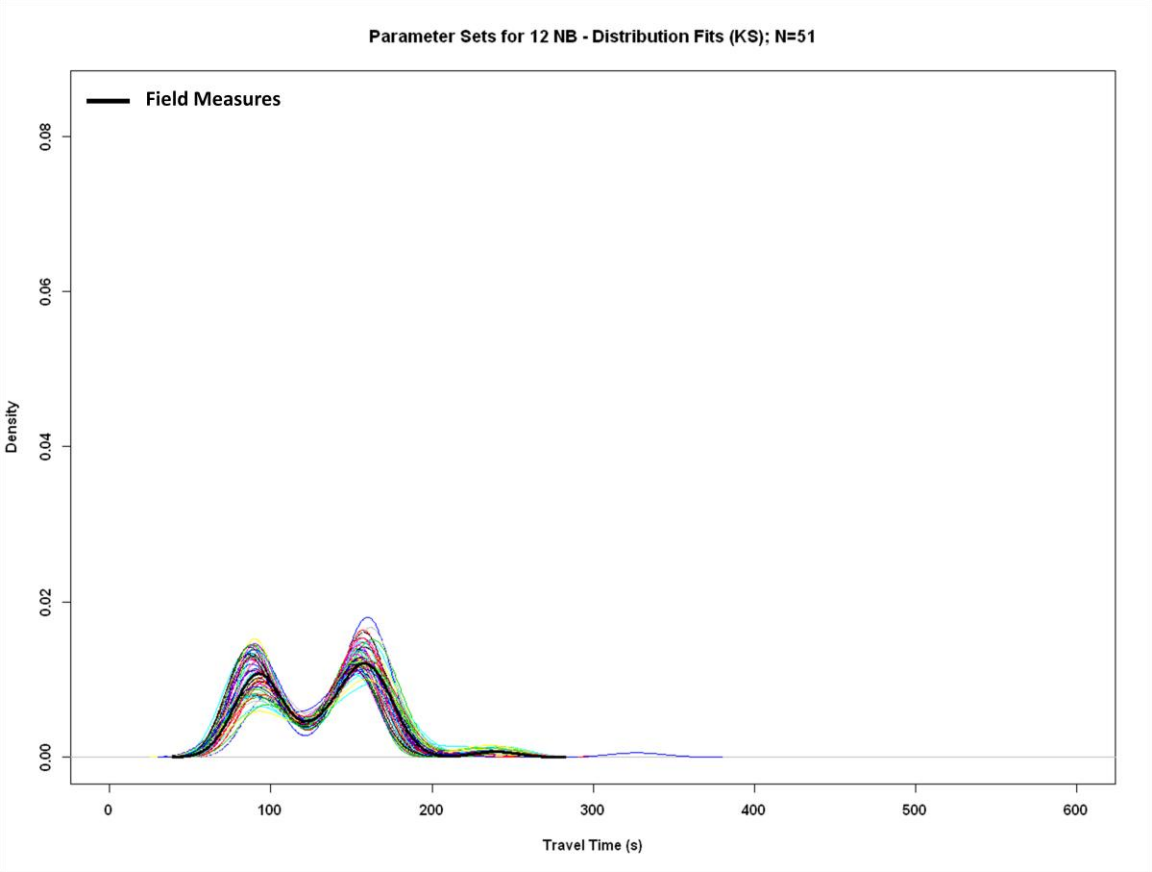


Figure 63 Models from KS test, $M_n|D$

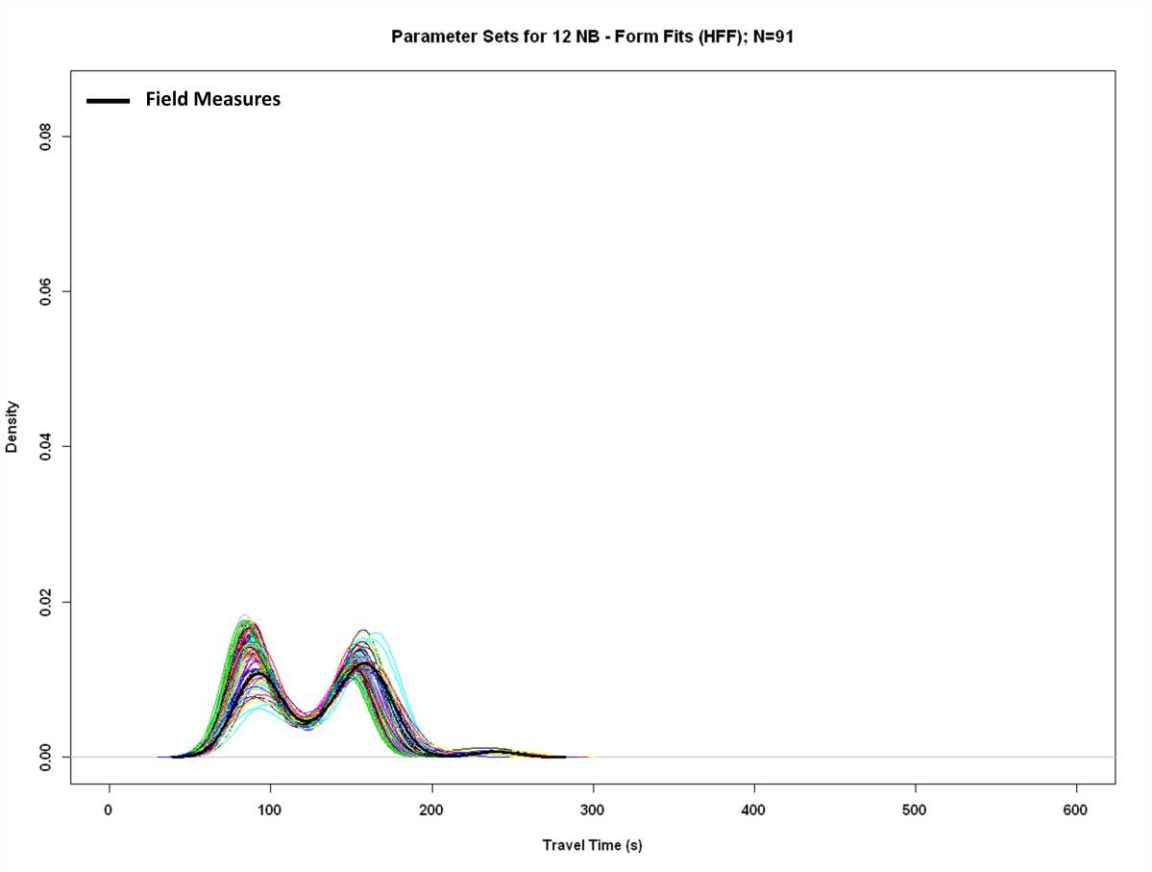


Figure 64 Models from HFF test $M_n|H$

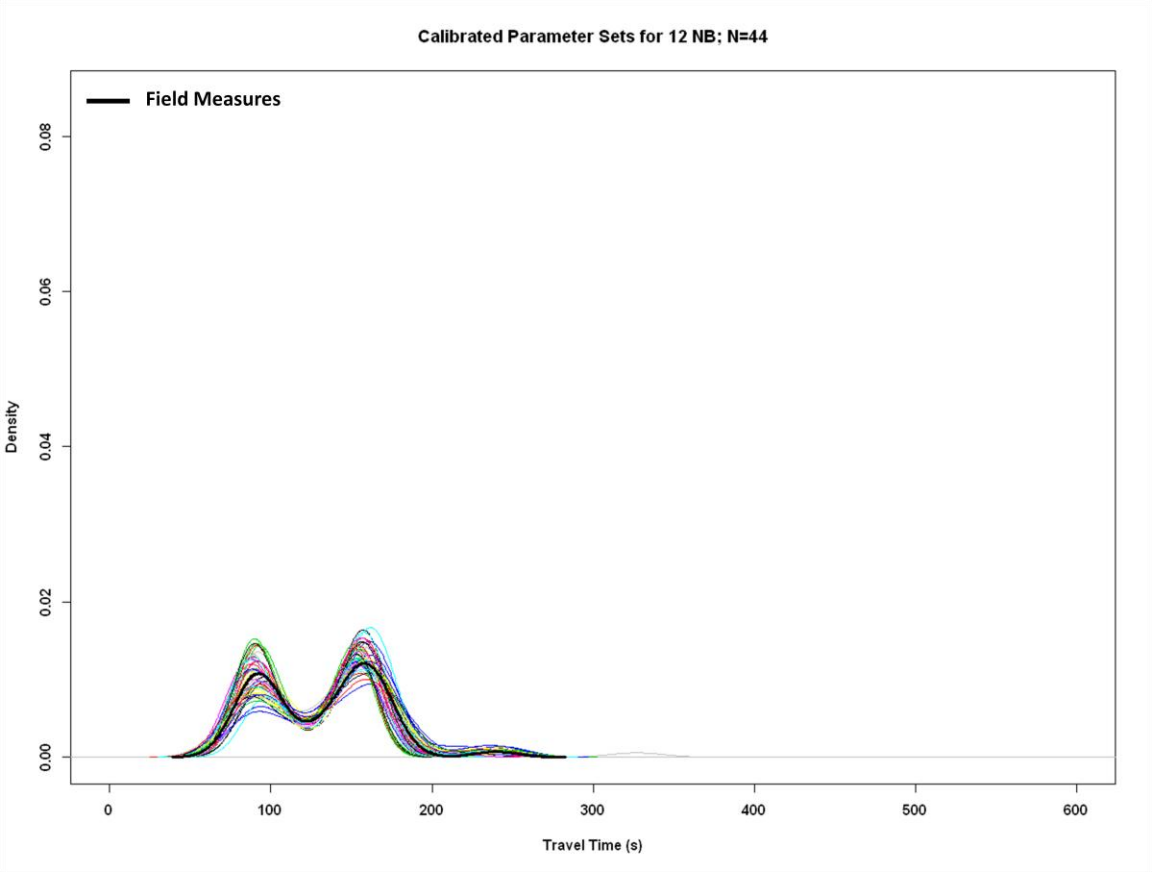


Figure 65 Final Set of Models – Noon Northbound

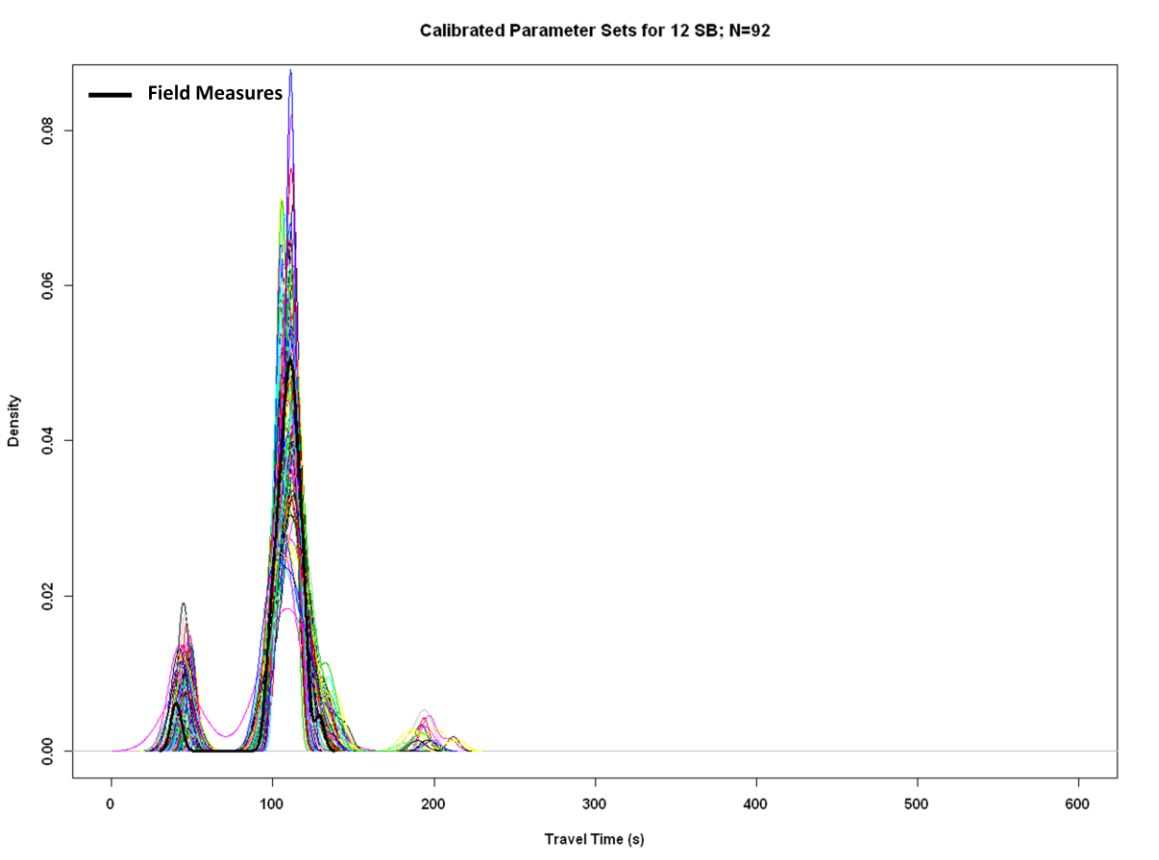


Figure 66 Final Set of Models – Noon Southbound

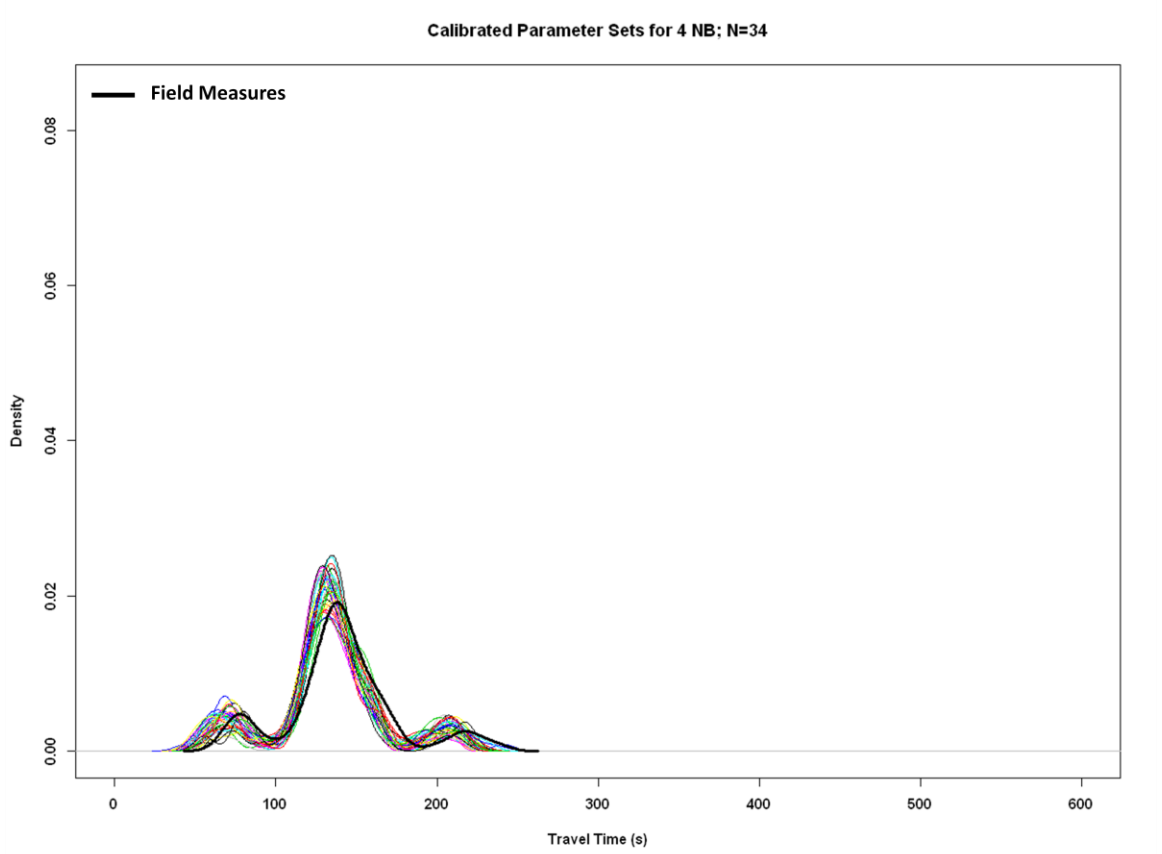


Figure 67 Final Set of Models – Evening Northbound

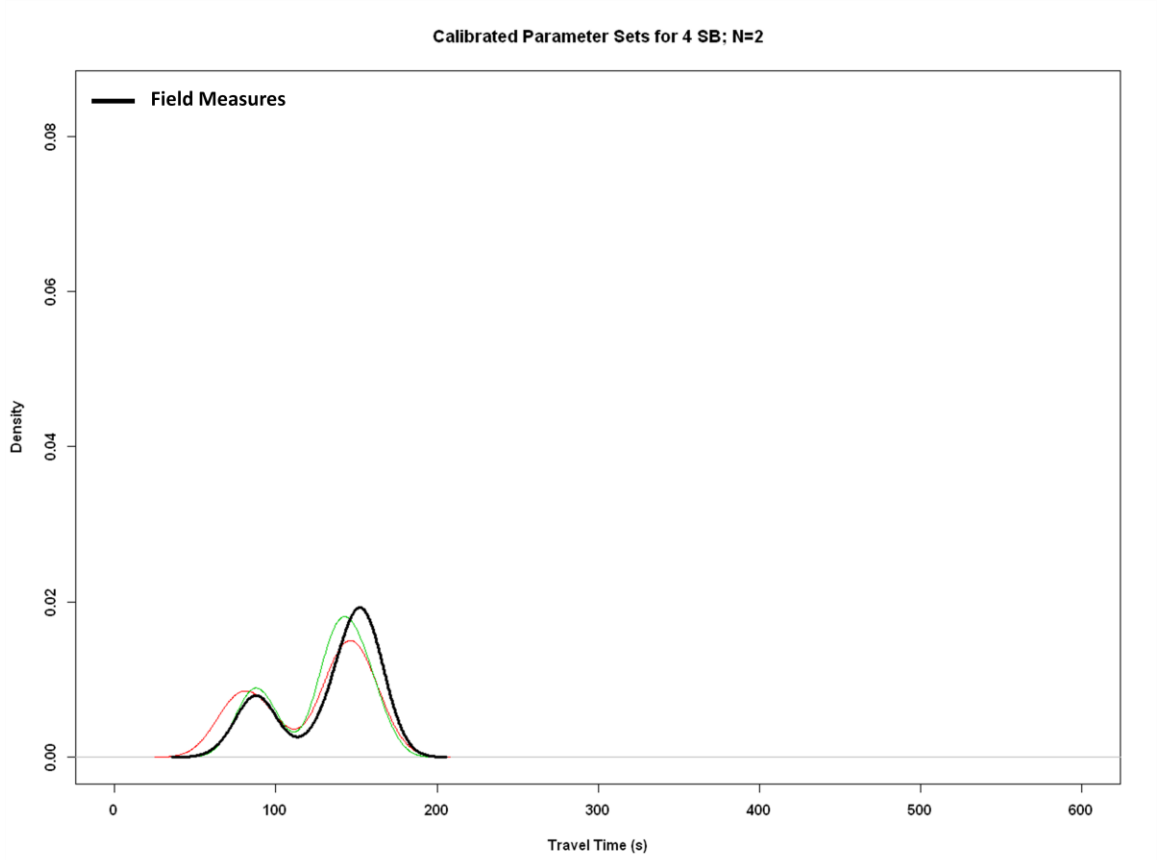


Figure 68 Final Set of Models – Evening Southbound

Table 21 Number of Parameter Set Remaining After the Application of Each Calibration Criteria

Criteria	12 NB	12 SB	4 NB	4 SB
	Number of Replicates			
Initial	1000	1000	1000	1000
Saturation Flow	159	159	159	159
Statistical Evaluation				
WMW	44	93	35	2
KS	51	101	35	2
HFF	91	122	100	158
Satisfied Criteria	44	92	34	2

5.2.2 Final Calibrated Parameter Set Selection

A final set of calibrated parameter sets must next be selected based on comparisons with four different field datasets – two periods with two travel directions. The preceding analysis resulted in a selection of parameter sets that are adequately calibrated for each time period – direction alternatives, $M_{f|12/NB}$, $M_{f|12/SB}$, $M_{f|4/NB}$, and $M_{f|4/SB}$, where the subscript represents time period and direction (i.e. $M_{f|12/NB}$ represents the set of adequately calibrated replicates for the 12PM, northbound traffic). Ideally, the parameter sets for each of these time periods and directions should be the same. However, the above analysis yielded a number of different calibrated parameter sets across periods and directions. This means, for each time period and direction, different parameter sets were able to produce a calibrated model. Table 22 presents the number of parameter sets that are the same for different time periods and travel direction.

Table 22 Number of Common Parameter Sets for each Approach and Time Period

	12-NB	**12-SB**	**4-NB**	**4-SB**
12-NB	44	43	29	1
12-SB	43	92	31	2
4-NB	29	31	34	2
4-SB	1	2	2	2
Common to all Periods and Directions	**1**			

There was only one parameter set that produce calibrated simulations for both period and travel direction. Figure 69 Figure 69- Figure 72 present travel time density plots from this single calibrated parameter set, as well as plots from the field data and the original VISSIM model with default parameter values. For comparison, Figure 73 and Figure 74 show the startup and saturation flow plots for the same parameter set.

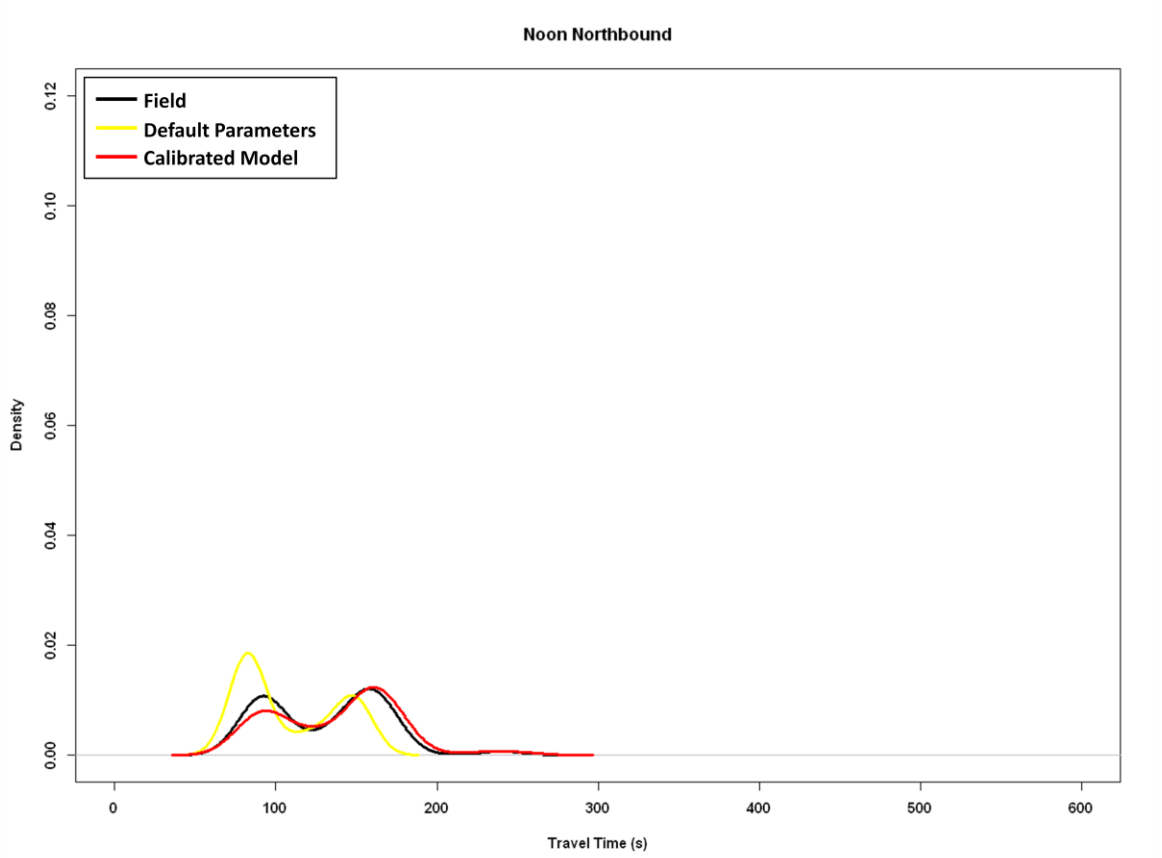


Figure 69 Travel Time Density Plots from the Calibrated VISSIM Model – Noon Northbound

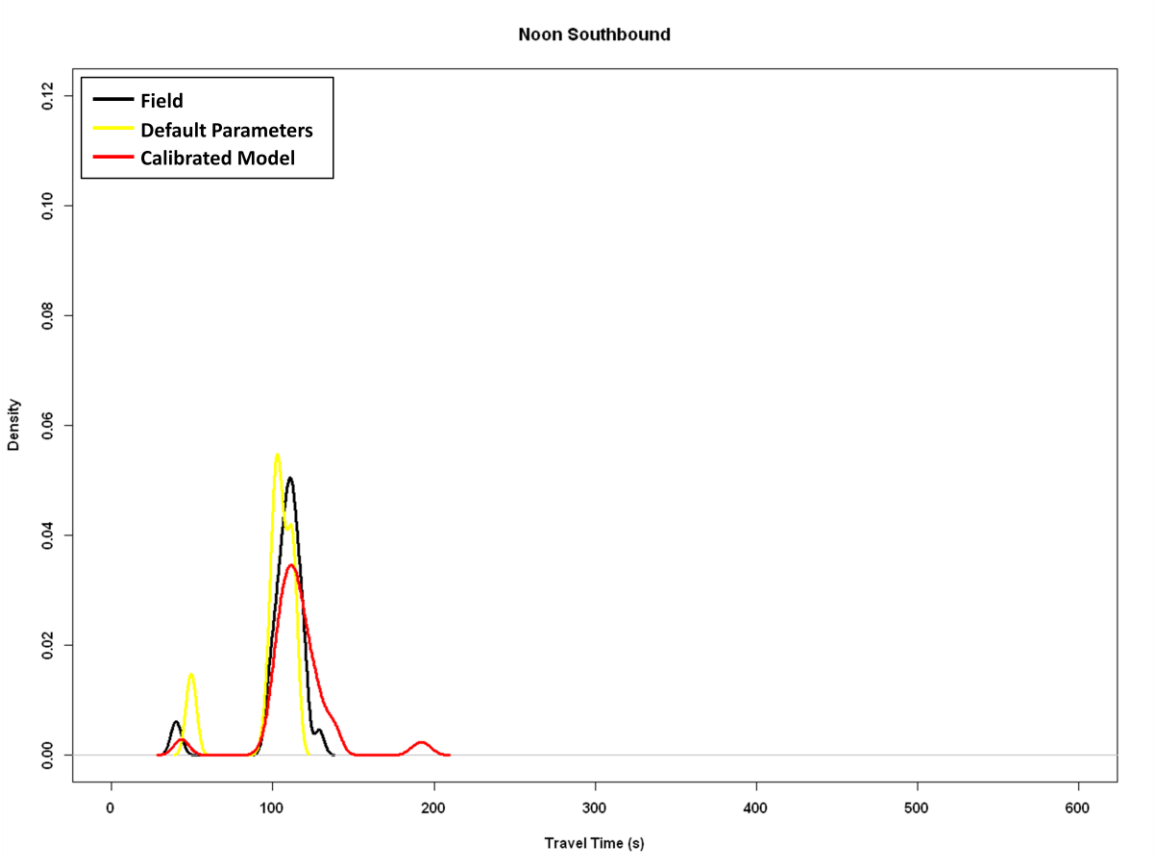


Figure 70 Travel Time Density Plots from the Calibrated VISSIM Model – Noon Southbound

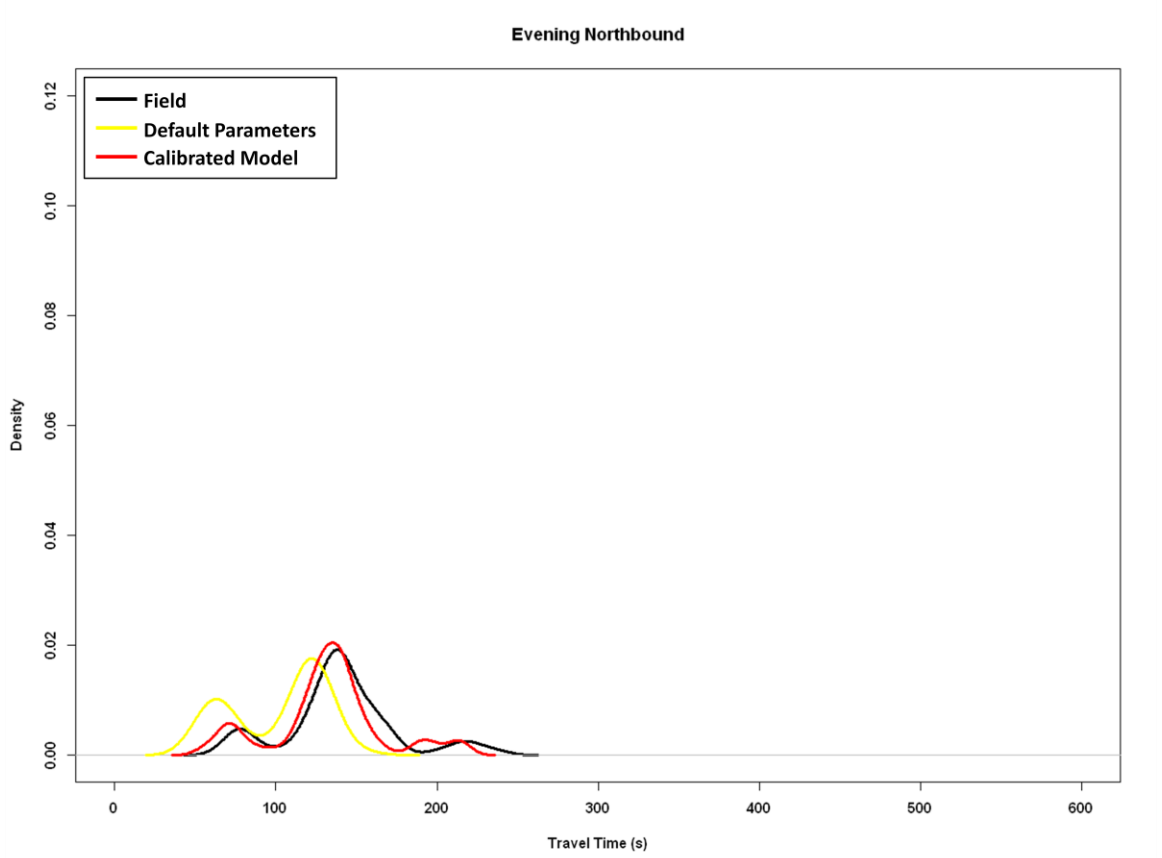


Figure 71 Travel Time Density Plots from the Calibrated VISSIM Model – Evening Northbound

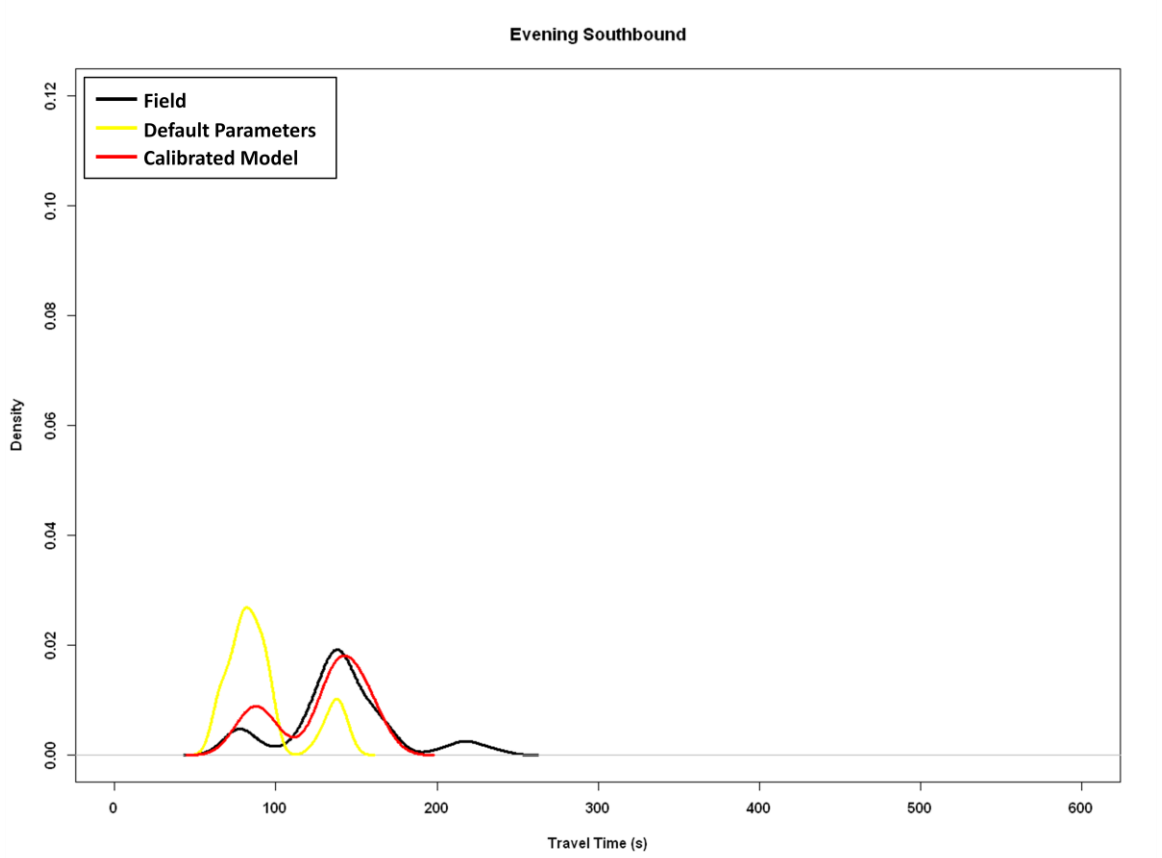


Figure 72 Travel Time Density Plots from the Calibrated VISSIM Model – Evening Southbound

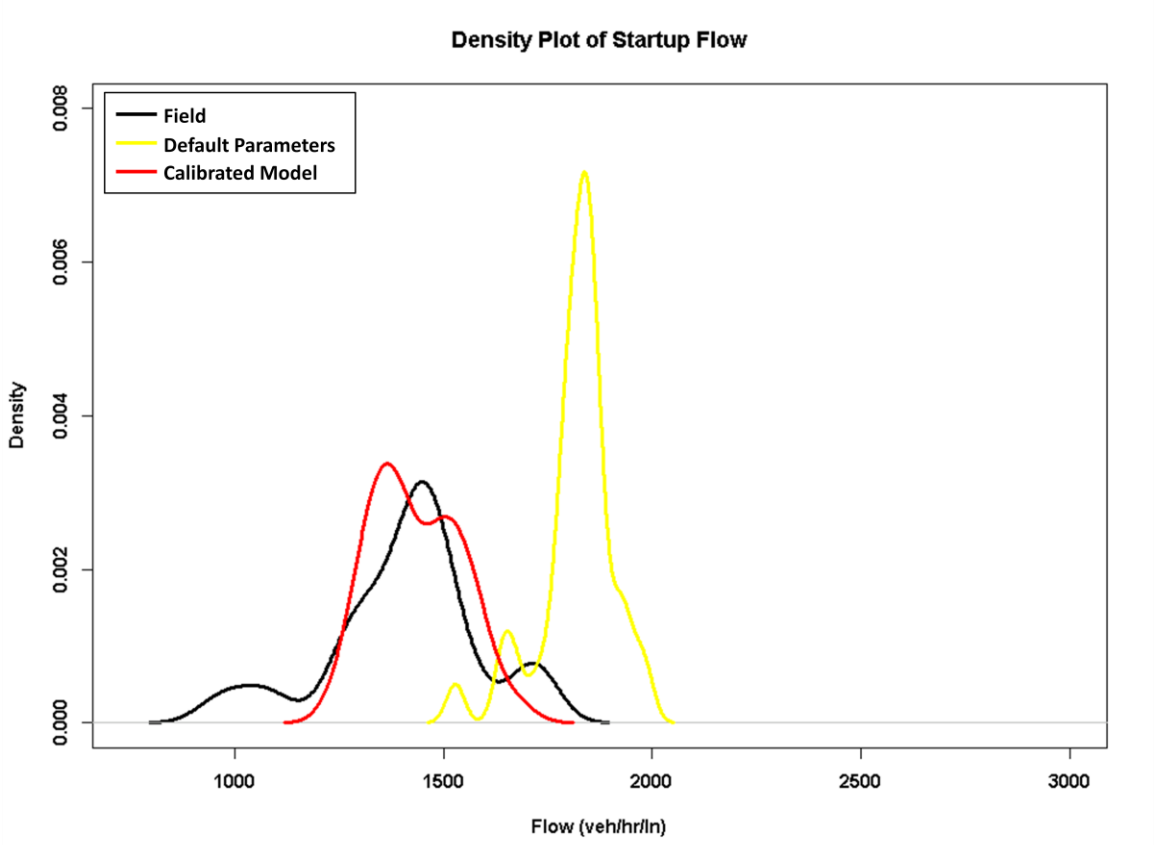


Figure 73 Startup Flow Density Plots from the Calibrated VISSIM Model

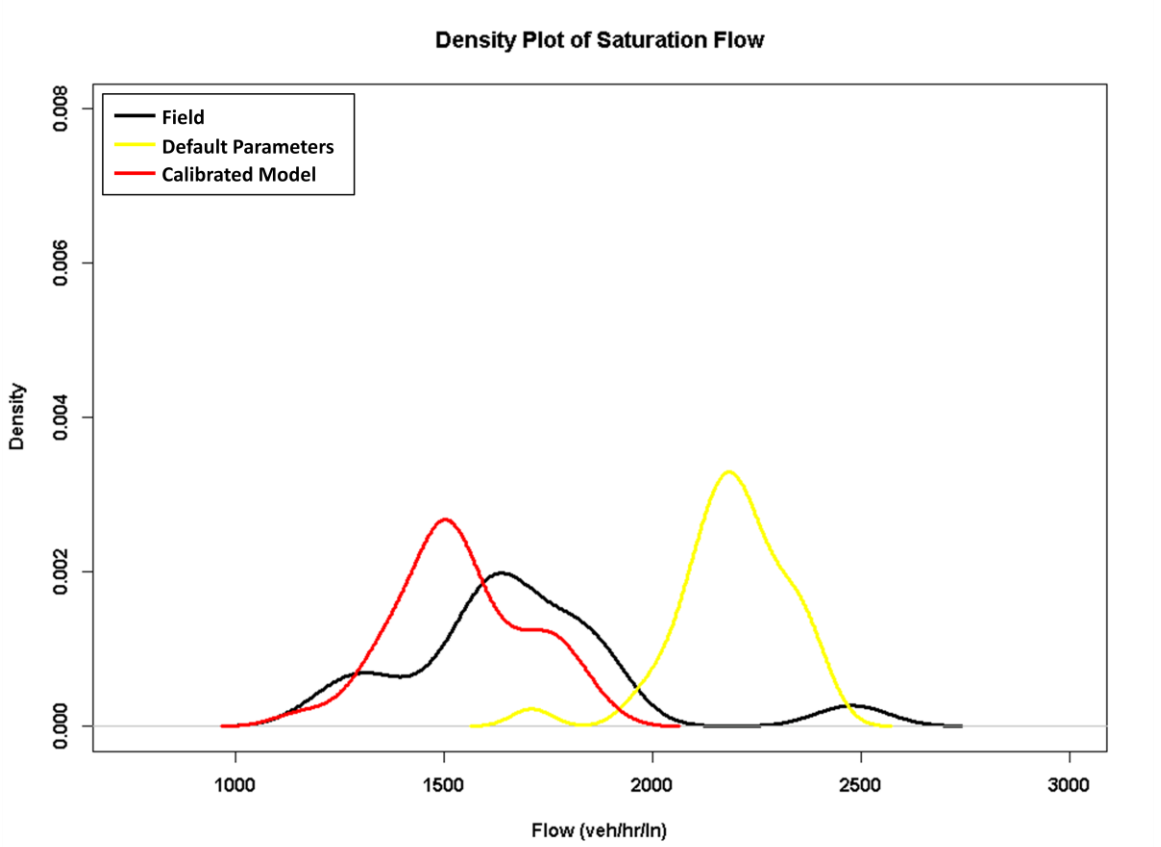


Figure 74 Saturation Flow Density Plots from the Calibrated VISSIM Model

Given the various sets of parameter values that produced calibrated simulation models, the next step is deciding which parameter set(s) should be used for real world applications. Two types of objectives are presented for accomplishing this objective.

The first objective is associated with planning purposes, which typically involves the evaluation and communication of macroscopic traffic information. If the intention of the model is to provide general information regarding traffic's current and future conditions, especially as it relates to the effects of certain strategies or policies, or changes in demand, it is recommended that any of the calibrated parameter sets may be used. In the case of the NGSIM study corridor, any one of the 95 parameter sets may be

used as they were deemed adequately calibrated for at least one period and direction. This recommend does allow for different time periods to utilize different parameter sets. However, where a parameter set(s) is found to satisfy all time periods and directions, as in this case, it is recommend that a parameter set from this group be used.

To obtain a more comprehensive view of traffic's performance, it is reasonable to perform an enhanced replicate study where not only the random seed is altered for each run but the parameter set may also change. In this instance, for each enhanced replicate, a parameter set is randomly selected from the family of calibrated sets. This approach parallels, and improves upon, the analysis of replicate runs in the traditional calibration process. Instead of only altering the random seed of a calibrated model, the values of the entire set of effective calibration parameters are changed. Such a change will provide greater insight into the variability of the simulated output. As a precautionary measure, it is recommended that throughout the trial analysis, the modeler should also monitor saturation flow measures in addition to the other performance measures, to ensure that a simulation's output remains valid.

The second category of objectives is more temporal in nature. This refers to models that are created to communicate microscopic, and macroscopic, traffic information in a time-base fashion, for instance in, or close, to real-time. For such purposes, it is advised that the calibrated replicates that best reflect traffic conditions during the analysis period be used. If traffic information about the noon period is requested, any number of the 136 parameter sets should be used to deliver the relevant information. Again, using more than one of these parameter sets is encouraged, making the information more robust.

One of the more immediate expansions of the effort is to have an array of *time-of-day* calibrated parameter sets which are intended to provide time-based traffic information for periods in which driver behavior maybe be deemed homogeneous. The use of these *time-of-day* based parameter sets is analogous to applying different signal timing plans, throughout the day, for an intersection. It is expected that there will be multiple calibrated parameter sets for each period. And similar to previous scenarios, these parameter sets should be used as a part of a replicate trial effort – providing a comprehensive view of anticipated performance measures. A further expansion of the *time-of-day* parameter set construct is the development of a real-time calibration algorithm. The goal of such an algorithm is to allow a real-time data driven simulation model to adapt to changes beyond those of the physical transportation network. This algorithm will be responsible for intelligently adjusting the effective calibration parameters in order for the simulated environment to continue accurately estimating measures – even as driver behavior changes.

5.3 Selecting values for effective calibration parameters

One of the opportunities that this advanced calibration process affords is the exploration of the values for each effective calibration parameter and its ability to produce a calibrated simulation. The objective in this section is to provide some guidance to the modeling community regarding how to choose and combine values for effective calibration parameters in order to increase the likelihood of producing an adequately calibrated model. To do this, the parameter values for all replicates will be compared to

those that produced an adequately calibrated simulation, for both periods and directions – set M_f, where:

$$M_f = M_{f|12/NB} \cup M_{f|12/SB} \cup M_{f|4/NB} \cup M_{f|4/SB} \qquad \textbf{(7)}$$

Figure 75, an inspired star (spider) plot [58] or parallel coordinate plot, demonstrates that the values for each calibration parameter and their various combinations were sufficiently covered by all the candidate parameter sets that were produced by the initial Monte Carlo process. The y-axis represents normalized parameter values so that they range from zero to one. The x-axis represents each of the effective parameters in addition to the "distance between cars" parameter, *d*. This parameter was not included in the original list of effective parameters as the modeler indirectly sets its value. The *d* parameter is a function of other parameters that are included in the set of effective calibration parameters (Equation 8). Each colored line in the figure below represents a single parameter set and its intersection with a vertical axis reflects the value of that particular parameter.

$$\boldsymbol{d = ax + (bx_{add} + bx_{mult} \times z) \times \sqrt{v}} \qquad \textbf{(8)}$$

where *v* is the vehicle speed (the mean desired speed was used a representative for each model) and *z*, which ranges between zero and one and is *N(0.5,0.15)*, was set to 0.5 [40]

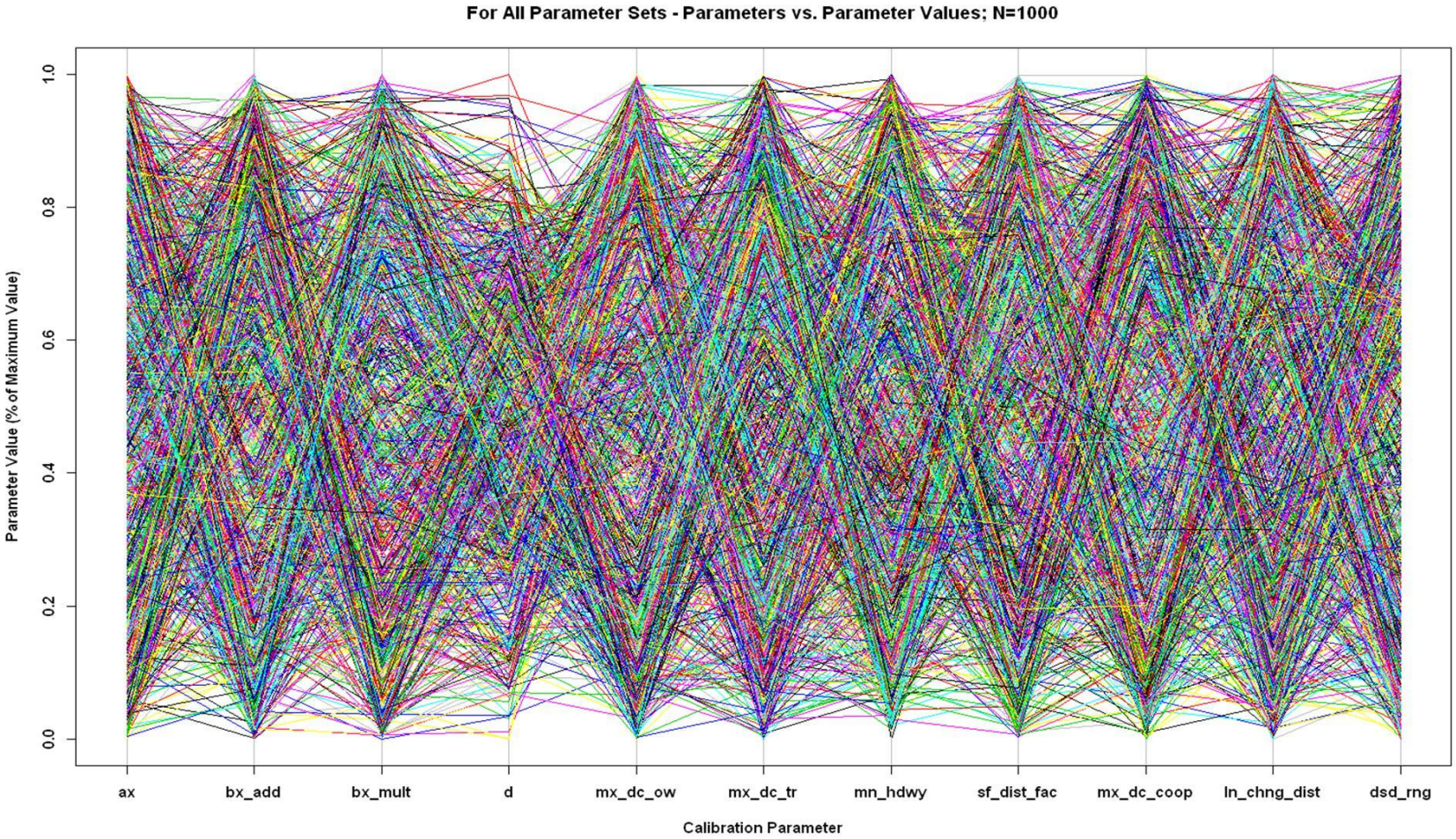


Figure 75 Candidate Calibration Parameter Sets and their Values

To help inform the selection of parameter values that increase the likelihood of producing a calibrated model, a similar diagram to the one above was created, Figure 76. The difference between Figure 75 and Figure 76 is that the latter only presents parameter sets from the set M_f for all time period and direction combinations. It was anticipated that Figure 76 presents distinct patterns in parameter values and these patterns will in turn inform modelers as to how to assign values to effective calibration parameters. However, distinct patterns were only observed for a few parameters. To further explore these patterns Figure 77 presents the density plots of the values for each parameter.

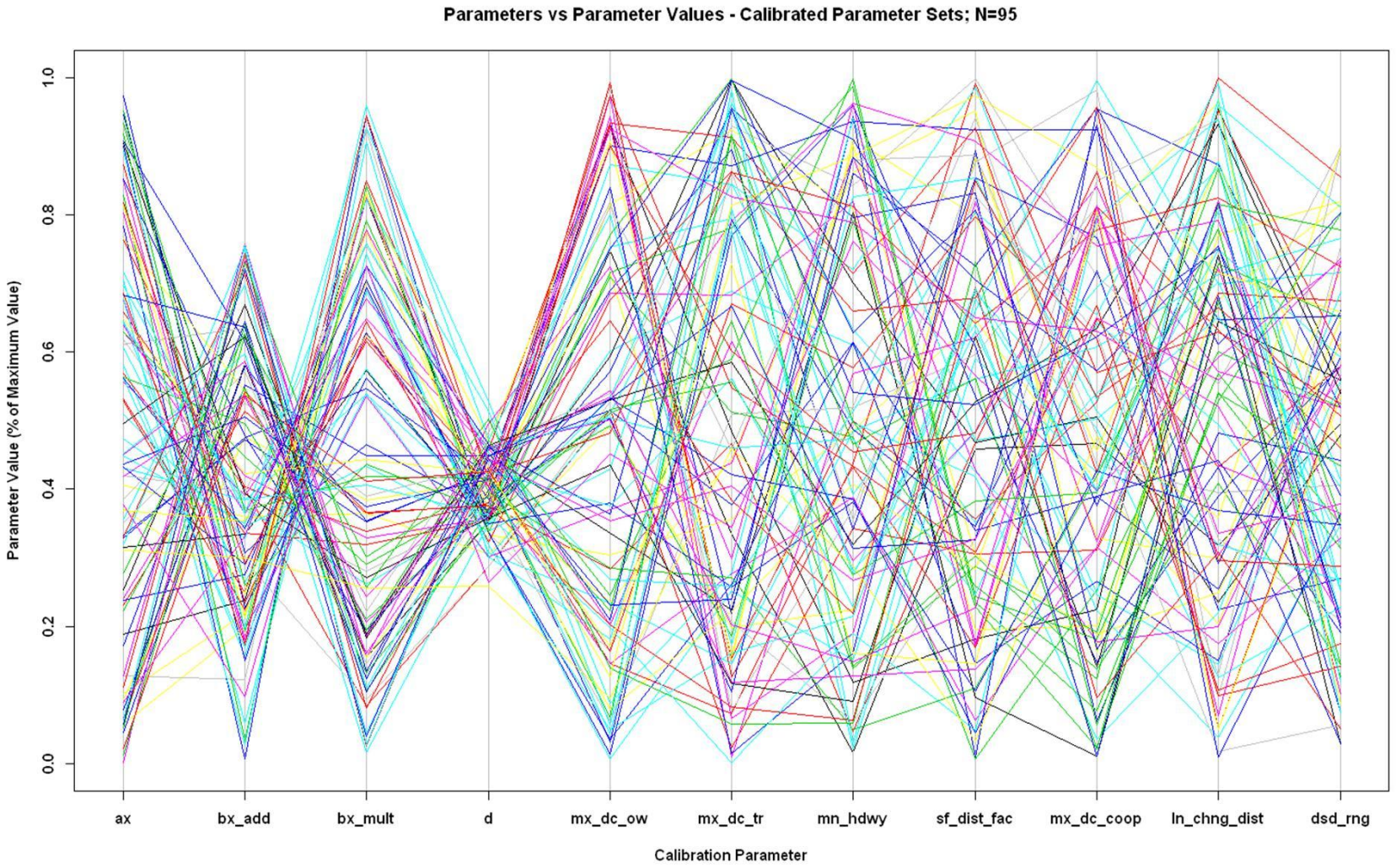


Figure 76 Calibrated Parameters Sets

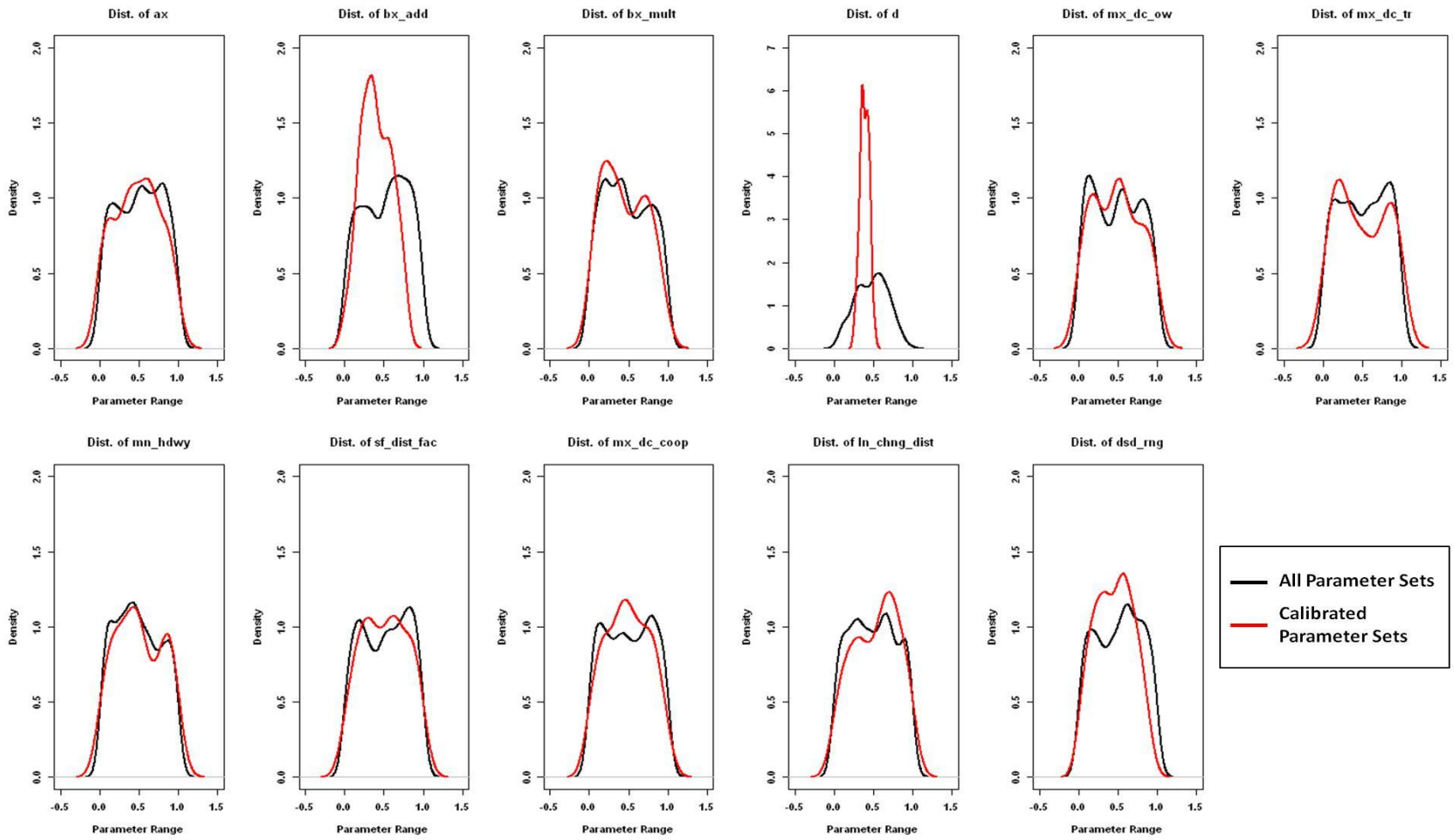


Figure 77 Density Plots of Parameter Values

Figure 76 and Figure 77 present strong patterns for parameters *ax*, *bx_add* *bx_mult* and *d*. Typically, for a given calibrated parameter set, as the value of *ax* approaches its maximum value, the value of *bx_add* approaches its minimum and *bx_mult* approaches its maximum value; and vice versa. The strongest pattern was observed in the values for *d*. All of calibrated parameter sets produced a *d* value between 26% and 53% of its maximum value. What this observation suggests is that the values for parameters, *ax*, *bx_add* and *bx_mult*, are acting in a combined manner to result in a *d* between 26% and 53% of its maximum value. With this knowledge parameter sets could be trimmed prior to running simulations that did not meet this constraint.

In the larger context of model calibration, the dominant patterns in *ax*, *bx_add* and *bx_mult* indicate that accurately estimating saturation flow is critical when developing a calibrated model. The VISSIM manual states these parameters are the main parameters that dictate flow [40]. Therefore, an accurate field estimate of saturation flow and a subsequent reasonableness range is imperative to establish in order to ensure that parameter sets are calibrated and that their results are suitable for real-world applications.

The density plot for the desired speed range (*spd_rng*) parameter in Figure 76 marginally suggests that narrow range of values are slightly more likely to produce a calibrated model. However, given that there are quite a number of calibrated parameter sets with *rnd_dsd* values at or near its maximum, modelers are encouraged to maintain a larger for this parameter. Based on results from this study a suggested range for this parameter is 15-20 mph. (See Section 5.3.1 for additional discussion regarding desired speed and desired speed range.) As for the other effective parameters, patterns in their assigned values for calibrated replicates are less distinguishable. Therefore, it is

seemingly just as likely for the default values of these parameters to produce calibrated replicates, as it would be if any other values were assigned. Thus, a future study could explore a reduced calibration effort in which only the four discussed parameters are considered and the other parameters are set to their default value. Based on the given results it is highly likely that this approach would result in calibrated models of nearly the same quality with a significantly reduced computational effort.

5.3.1 Desired Speed and Desired Speed Range of Calibrated Replicates

The following section will present the comparison of mean desired speeds and their respective ranges for the set of calibrated parameter sets. This comparison further explores the suggested width of the desired speed range parameter (15-20 mph). In addition, this comparison also follows the recommendation presented in Section 5.1.1.1, by examining the mean desired speed and the associated range for the calibrated set of parameters, to ensure practical desired speeds.

Figure 78 presents two plots that compare mean desired speed and desired speed range. The plot on the left depicts the comparison for all parameter sets, while the one on the right depicts the comparison for only calibrated parameter sets. This figure illustrates that although some of the initial parameter sets included means and ranges that may produce impractical speeds, a significant number of those parameter sets failed to meet the calibration criteria. Therefore, the achievable speeds from the selected parameters are probable and consistent with desired speed estimates from the field.

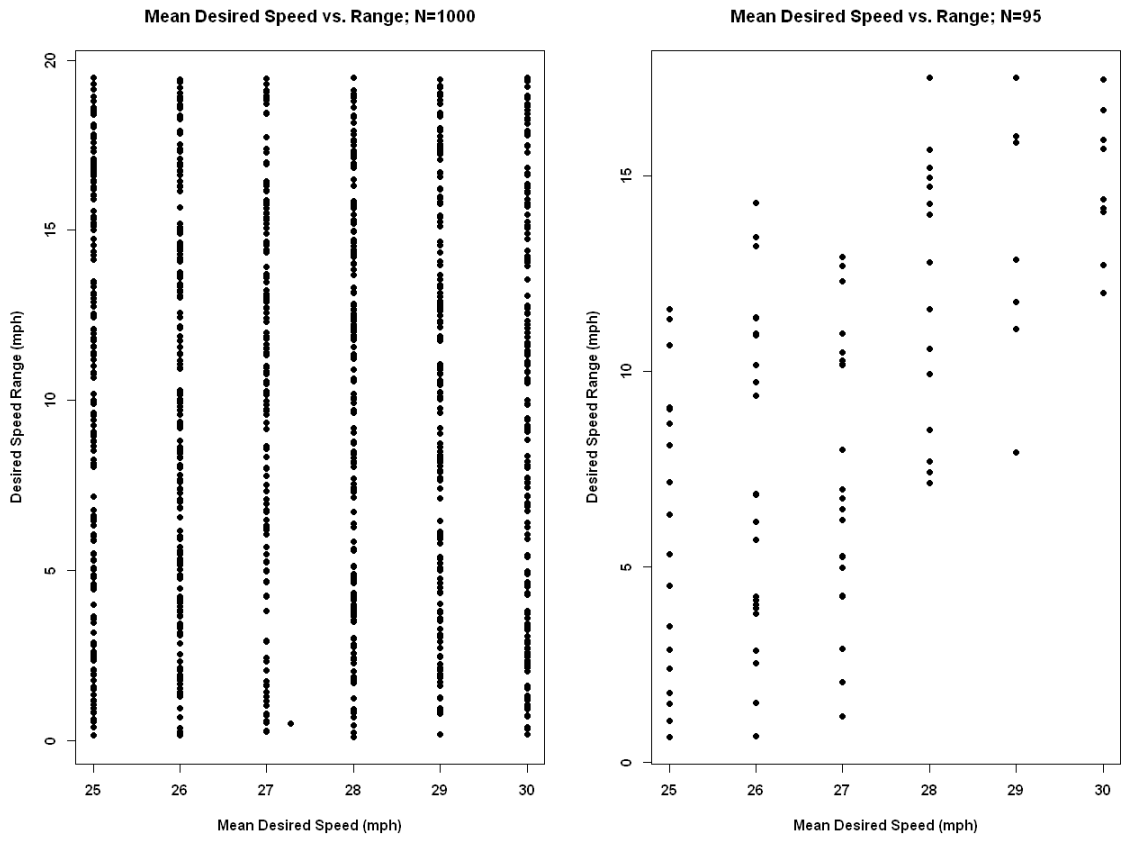


Figure 78 Mean Desired Speeds and the Respective Ranges

Table 23 presents the minimum and maximum speeds that are achievable from the calibrated parameter sets.

Table 23 Mean Desired Speed and their Ranges of Calibrated Parameter Sets

Mean Desired Speed (mph)	Desired Speed Range (mph)		Desired Speed (mph)	
	Min	Max	Min	Max
25	0.6	11.5	13.5	36.5
26	0.7	14.3	11.7	40.3
27	1.2	12.9	14.1	39.9
28	7.2	17.5	10.5	45.5
29	7.9	17.5	11.5	46.5
30	12.0	17.5	12.5	47.5

When comparing Table 14 and Table 23, and examining the results from Figure 78 it may be conclude that the desired speed range parameter should have a maximum value between 10 or 15 mph. However, the effectiveness of this narrower range is dependent on the accuracy with which the field mean desired speed is estimated. Referring to Table 14, the minimum and maximum field estimates of desired speed were all within 11 mph of their respective means. This suggests that a maximum value of 10 and 15 mph may be appropriate for the desired speed range parameter in this case. Figure 78 also supports this inference as approximately 90% of the calibrated parameter sets had a desired speed range less than or equal to 15 mph (and 57% of the ranges were less than or equal to 10 mph).

There were desire speed ranges that were greater than 15 mph. However, these corresponded to mean desired speeds that were larger than the average desired speed of the study corridor. These larger speed ranges were therefore attempting to counteract the overestimation of desired mean speeds to allow a parameters set to satisfy the calibration criteria. Towards this end, it recommended that modelers take great care in estimating the desired speed of their study network. After an accurate estimate is obtained, a desired speed range of 10 mph is sufficient. However, a more conservative range of 15 mph is recommended.

5.4 Advanced Model Calibration – Summary

The preceding sections presented an advanced calibration procedure. This procedure was used to calibrate the VISSIM model that was part of the NGSIM pseudo field test in Chapter 4. This procedure used a set of ten effective calibration parameters,

whose values were assigned in a Monte Carlo fashion. These ten calibration parameters were deemed significant according to the sensitivity analysis research carried out by Miller 2009 [46] . At the foundation of this research is the evaluation of the values for each of VISSIM's relevant calibration parameters and the observation of their effects on travel time. Although these parameter were selected based on travel time, and the above procedure does evaluate parameter sets against startup and saturation flow, the final set of parameters were crossed checked against the VISSIM manual to ensure that parameters that do affect startup and saturation flow were included. In the future, a separate sensitivity analysis should be conducted on the initial set of parameters utilizing startup and saturation flow to ensure that no significant parameters were overlooked by a travel time based sensitivity analysis.

A Monte Carlo approach was then employed to generate 1000 potential parameter sets for calibrating the corridor VISSIM model. These parameter set simulations were evaluated against a robust set of calibration criteria to determine which were calibrated. There are two parts to the calibration parameter value selection process: 1) evaluation of startup and saturation flow and 2) statistical evaluation of travel time distributions. The parameter sets that satisfy both these criteria are considered as adequately calibrated. This calibration procedure is summarized in Figure 79.

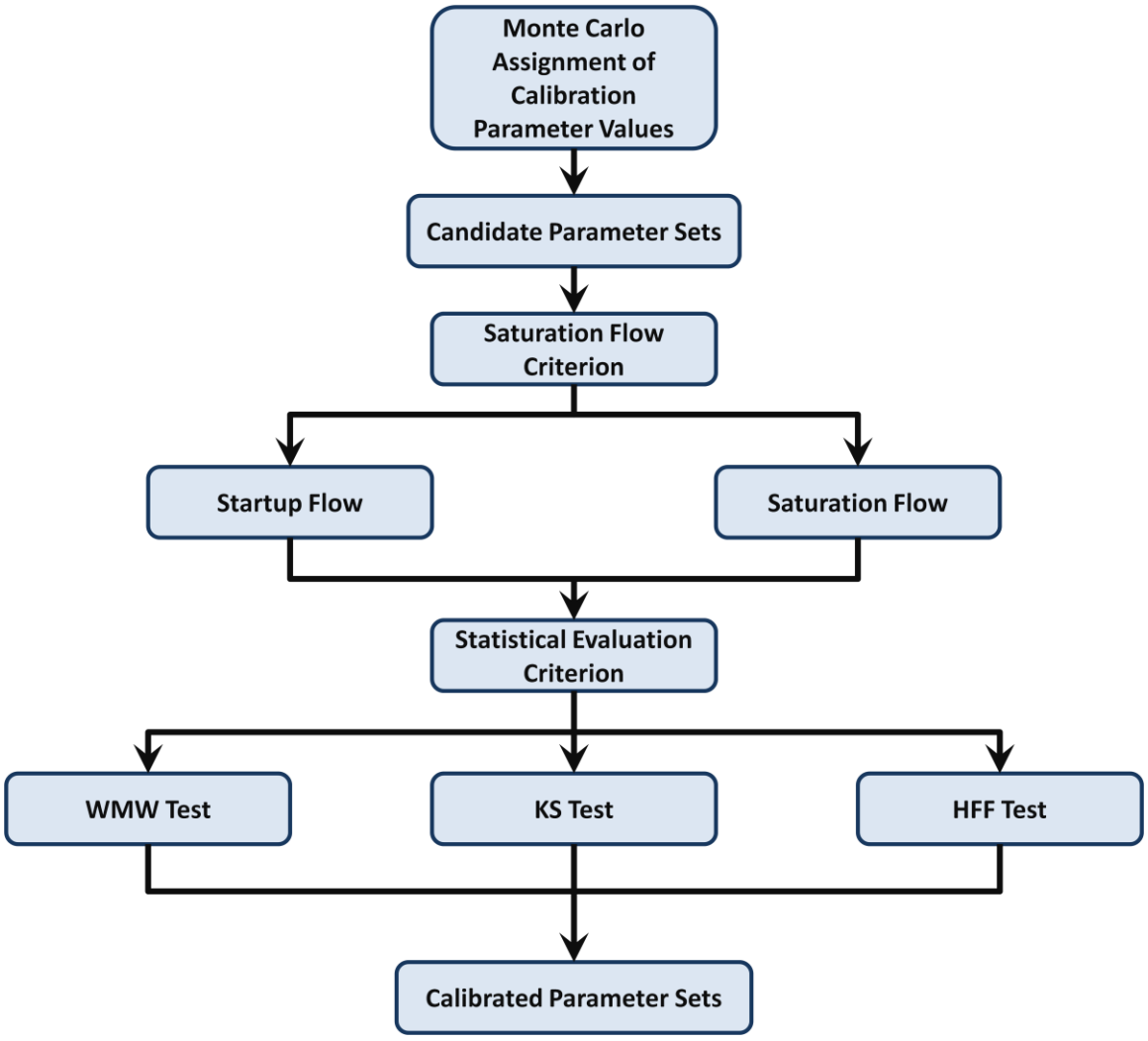


Figure 79 The Advanced Calibration Procedure

After executing the above procedure, it is likely that more than one parameter set will be determined to provide an calibrated model. To determine which parameter set(s) to use, the purpose of the use of the model should be defined. The above work delineates two potential purpose types – planning and temporal. For planning purposes any number of the calibrated parameter set may be acceptable while for temporal purposes, only those that model the traffic conditions pertaining to the period being evaluated should be used.

The structure of this calibration procedure provided insight into the assignment of parameter values in order to increase the likelihood of producing a calibrated parameter set. Inferences regarding parameter value assignment were made about four of the ten effective calibration parameters. The value of parameters *ax*, *bx_add*, and *bx_mult*, which determine flow, seem to interact to produce a *d* value that is between 26% and 53% of its maximum value. In the later discussions regarding the value for the desired speed range parameter a conservative value of 15 mph is recommended. However the modeler should be a aware that a value of 10 mph may also be used but the estimation of mean desired speed is encouraged to be as accurate as possible. As for the values of other parameters, evidence seems to indicate that default values are no less likely to produce a calibrated model than any other values. Therefore, for future implementation of a more efficient calibration process it is recommended that the parameters set for calibration be limited to only those found as critical in this process, potentially significantly reducing the computational time required to develop a reasonably calibrated model.

CHAPTER 6

FUTURE RESEARCH

The previous chapters demonstrated the feasibility an online, data-driven simulation tool to estimate arterial performance measures in real-time. However, to facilitate successful field implementation there is additional research needed to improve the method's accuracy and the robustness. Five opportunities have been identified by which to achieve this: 1) improve the accuracy of vehicular volumes entering the study network, 2) conduct real-time estimations of turning movement distributions, 3) synchronize field and model traffic signal phases, 4) real-time calibration and 5) modeling boundary conditions. It is anticipated that these actions will lead to more accurate and reliable performance measure estimates. The following sections will detail challenges and possible solutions to these opportunities.

6.1 Vehicular-Volume Accuracy

Errors associated with vehicular volume were prevalent throughout the third field experiment. When observing vehicles entering a camera's detection zone, one of three events occurred: 1) the zone was appropriately triggered and as a vehicle exited the zone a car was generated in VISSIM; or 2) a single vehicle (usually a larger vehicle) triggered two detection zones and VISSIM generated two vehicles; or 3) a vehicle failed to trigger the detection zone and no vehicle was generated by VISSIM.

An accurate estimate of the number of vehicles entering the network is imperative given the intricate interrelationships among, volume, speed and density along arterials. The relationship among these three parameters is well defined for freeway traffic [59].

However, for arterials, their relationship is still being formulated and therefore, well-defined analytical means are unavailable to make adjustments to account for incorrect volume inputs. It is therefore critical to have accurate estimates of vehicle volumes. These estimates will in turn increase the fidelity with which the simulation environment reflects the field.

To better estimate the number of vehicles entering the study corridor a three-pronged approach has been indentified to address this issue. First, the detector manufacturer's guidelines should be consulted to ensure optimal settings and detector placement, which will in turn reduce incorrect vehicle detection. It is hypothesized that these optimal settings will reduce the frequency of incorrect detector triggers. Second, an extensive review of previous work related to improving detector accuracy should be conducted. This review will examine the use of inductive loop detectors and other detector technologies, and the techniques employed to address their associated erroneous detections may be extrapolated to suit the needs of this research effort. Third, other detector information will be used to verify the presence of vehicle. For instance, headway and/or occupancy information can help to decipher whether or not a vehicle entered the detection zone. This three-pronged approach will potentially increase the accuracy of the number of vehicles entering the network, making the overall real-time data-driven methodology more reliable and robust.

6.2 Turning Movement Distribution

Accurate turning movement distributions are also needed to provide accurate real-time estimates of performance measures. The study of estimating turning movement

portions is similar to the development of models that dynamically estimate origin-destination flows in a small network [35]. Where vehicles utilize separate lanes for turning movements (i.e. turn only lanes) it is possible to use current detection technology to estimate movements. However, where multiple movements share a lane, it is not readily possible to determine the movement of a vehicle through the intersection once the vehicle departs the detector.

Henry Liu et al. (2009) [35] and Chang and Tao (1998) [60] presented a summary of some the more notable works involving the estimation of turning movement proportions. Included in these works are the authors' own models for determining the proportion of drivers that execute the various turning maneuvers at an intersection. These models use short-term detector counts and vehicle groups, according signal phases, to determine turning movement proportions. Although these models, along with others, provide fairly accurate proportions they have yet to demonstrate their ability to be readily incorporated in a real-time framework. Even though these models have not been validated in a real-time environment, each possess elements that could be used to inform the development of an algorithm that accurately estimates turning movement proportions in real-time.

The approach that is currently being considered uses intersection travel times and a series of related density functions to determine a vehicle's path through an intersection. Each approach has a series of travel time density functions that corresponds to the turning movement at an intersection. Therefore, as a vehicle travels through the intersection, its travel time is inserted into a density function and its "likely" path determined. Despites preliminary success of this approach there are two caveats that may limit its

implementation. First, the method requires upstream and downstream detectors for each approach. Such detector configurations are uncommon but feasible depending on the implemented roadway detection technology. Second, this method requires a lag of a few seconds as vehicles need to traverse the intersection before the simulation can assign a path through the intersection to a vehicle. As a result, the term "real-time" may have to be redefined and therefore other portions of the framework should be adjusted to accommodate this lag.

These current efforts assume that a detection technology is not available that would readily assign the path of a vehicle through and intersection. New technologies should also be sought that allow the vehicles path to be directly measured from the field, eliminating the need for estimations.

6.3 Signal Synchronization

During the third field experiment there were instances when the field signal indications were different from their representation in the simulated environment. Given the direct relationship between signal operations and arterial performance measures, it is imperative that field and simulated signal states are synchronized.

To establish synchronization of the traffic signals in the two environments, prior efforts should be consulted and, if applicable, implemented or used as foundation of a solution for this research effort. One effort that should be closely examined is presented by Ban et al. (2009) [61]. Ban et al. estimated signal timing plans using piecewise linear intersection delay curves and a two-step least square estimation algorithm. Although this method may not be suited for the current real-time approach, studying its application of a

virtual trip line technique, translated signal phase timing, and vehicle penetration rates will provide insights as to how to achieve and automate signal synchronization.

For the third field test, a simplified two-phase signal synchronization method was developed and implemented. This two-phase method first input the appropriate signal timing plans into the simulated environment. The simulated signals were forced into synchronization by controlling the simulation clock at the initiation of the experiment. However, this method is not sufficiently robust for a full scale, permanent implementation, as signal synchronization was only temporary and also fails under actuated signal control environments. It was seen in the field experiments that synchronization was lost under two main scenarios: 1) when a gap-out condition is met in the field but not in the simulated environment and 2) when the duration of the experiment goes beyond the single time-of-day timing plan that is currently implemented in VISSIM.

When the signals become unsynchronized under the first condition, they tend to return to a state of synchronization, in part due to fixed background cycle lengths. However, under the second condition, synchronization is lost when the underlying simulated timing plan differs from the field plan. When time-of-day plans change, it is common for most, if not all the parameters of a timing plan to be adjusted. Once parameters such as cycle and phase lengths change, "re-synchronization" is highly unlikely unless the simulation's corresponding time-of-day is simultaneously activated. To prevent the second condition, the current method only simulates a during a single time of day plan.

To develop a more robust signal synchronization plan, a means to stream signal states into the simulation should be investigated. Two potential approaches to stream

signal states include 1) the signal controllers transmit the necessary information via wired or wireless communications or 2) the use of wireless camera within the cabinet to monitor the load switch indication and thus signal indications. In the first approach, a potential limitation exists in that the required frequency with which data needs to be streamed, one hertz, may exceed the capabilities of today's controllers. Although streaming signal data at a lower frequency is possible, a frequency of 1 Hz is preferred as lower frequencies have the potential to reduce the method's ability to achieve and maintain signal synchronization. Nonetheless, it is worth exploring hardware and/or software patches that can extend the controller's streaming capabilities.

The other approach being contemplated is the use of wireless cameras inside the controller to monitor and transmit switch pack indications. The transmitted video will be processed and phase indications extracted and used to drive the signals in the simulated environment. Through this approach, signal synchronization will be achieved and maintained, regardless of gap-out conditions or changes in time-of-day plans.

6.4 Calibration

The calibration process that was developed and implemented throughout this method was performed after the field test. Although calibrating the simulation models offline was successful, future efforts should seek to bring this process online. Real-time calibration will enable the proposed system to become more robust as it will be able to adapt to non-facility changes such as driver behavior.

In addition, the next round of calibration efforts should also attempt to accurately model pedestrian-vehicle interactions. The need to represent pedestrian-vehicle behavior

clearly arose from observing changes in driver behavior and performance measures due to pedestrian activity from field experiment #3. Therefore, for the simulation environment to continue to accurately estimate performance measures, pedestrian-vehicle behavior needs to be represented in the simulation environment and as part of the calibration effort. Suh et al. (2012) [62] introduced a method to effectively model pedestrian-vehicle interactions in VISSIM. Next steps in the real-time model development should seek to build on this method to include modeling pedestrian activity and to have such activities be a part of the real-time calibration process.

6.5 Boundary Conditions

The challenge of real-time simulation is to mirror dynamic traffic conditions in real-time. As part of this effort, it was observed that the simulation model was capable of reflecting congested conditions when the cause of the congestion (i.e. bottleneck) resided within the simulation boundaries. However, if the source of congestion was outside the simulation boundaries and spilled back into the simulation region this was not captured. For example, if an intersection outside of simulated region resulted in queues blocking an upstream intersection within the simulation boundary this would not be reflected. The underlying challenge is the development of an ability to restrict flow on simulation exit links such that the blockage due to downstream congestion is reflected. To date VISSIM does not have an immediate solution to this challenge. Therefore, future efforts will be aimed at dynamically controlling the flow rate at nominally unrestricted exit points in real-time.

6.6 Next Steps

This effort has utilized detectors and equipment specifically installed for the real-time simulation experiments. This installation allowed for highly detailed information to be streamed in real-time. The next phase of the research should seek to implement a real-time simulation on an arterial corridor utilizing data streams with a potentially lower fidelity.

The real-time research effort should also continue to explore the issues discussed in Sections 6.1 through 6.5, i.e. entering vehicle volume accuracy, determination of real-time turning movements, synchronization of field and simulated signals, online calibration, and downstream congestion influencing boundary conditions. Improvements in each of these areas will improve the performance of overall real-time simulation system. However, the next phase of the project could eliminate several of these issues through targeted selection of the next site, allowing for a more focused effort.

In addition, a larger field experiment would also need to explore potential communication challenges. The current test bed had the benefit of utilization of the campus fiber network. To be successful, the next implementation should investigate the communications between the detectors and the data processing center and communication between the processed data node and the simulation in a location without this benefit.

CHAPTER 7

CONCLUSION AND CONTRIBUTION

7.1 Closing Remarks

A wide variety of advanced technological tools have been implemented throughout the nation's transportation network to increase its efficiency. This research project explored the feasibility of integrating real-time detector data streams with an arterial simulation to support an arterial performance monitoring system. The information from this system is aimed at increasing network efficiency by enabling more informed decisions regarding network usage and management.

To begin exploring the feasibility of this approach, a simulated test bed was created. This test bed was comprised of two simulation instances. One instance represented the field while the other a "simulated" environment being driven by detector data streams. The results from this test-bed illustrated that the underlying approach being considered could be successful in estimating performance measures in real-time. Two field experiments were subsequently conducted, building on the success from the simulated test-bed.

The first field experiment involved the use of temporary detectors streaming PVR (per-vehicle-records) data directly into a VISSIM model of the study area. The results from this experiment revealed that the performance measures of the simulated environment and the field were similar. However, several limitations were noted, the two most dominant being the lack of a well calibrated VISSIM model of the study area, and the inability to have the signals in VISSIM synchronized with those in the field. This experiment underscored the feasibility of the method being explored despite the few

observed discrepancies. A second field experiment was conducted that sought to address some of the limitations that were encountered in the first field experiment and improve the simulations ability to estimate field performance measures. The second round of results demonstrated an improvement in the accuracy of the simulated estimates. Despite the success of these field experiments, there were still discrepancies between field and simulated estimates.

These discrepancies were attributed to a number of sources. These sources included: a) detector errors at simulation boundary detectors resulting in volume discrepancies between the simulation and field; b) differences between individual vehicle turning movements in the field and simulated environment; c) challenges in the synchronization of field and simulated signal indications; d) model calibration, and e) downstream congestion influencing simulations boundary conditions.

To test the proposed real-time approach in an environment that eliminated or significantly reduced the effects of the above sources of discrepancies, a "pseudo" real-time field test was undertaken. The test was conducted using data from the FHWA Next Generation Simulation (NGSIM) program. This high quality, high-resolution dataset was used to create a pseudo real-time data stream, which was then used to drive a simulation. This application of the proposed methodology provided an estimate of the ability of the system given ideal (i.e. no errors in the detector stream and no congestion spillback into the network) field conditions. While the accuracy in performance measure estimations improved, there were still noticeable discrepancies. Upon examination of these discrepancies, it became apparent that they may be addressed through additional calibration of the VISSIM model of the NGSIM study area.

Thus, an advanced, robust calibration procedure was developed and applied to the NGSIM model. This advanced calibration process used a list of known significant calibration parameters, along with a two-part criteria process, to determine adequately calibrated models. The first criteria ensured that the simulated saturation flow estimates were within a predefined range of reasonable field values. The second part of the criteria involves a series of non-parametric statistical analyses to evaluate the accuracy of travel time measures. After performing this advance calibration procedure, performance measure estimates were further improved.

This research effort has demonstrated that this data-driven, microscopic simulation technique potentially provides an opportunity to determine high fidelity arterial performance metrics in real-time. There are still challenges associated with this approach, especially as it relates to wide spread implementation. However, this initial effort developed techniques for addressing many of the challenges of real-time simulation, identified future challenges that remain to be addressed, and created a foundation upon which future implementations of real-time arterial simulation may be based

7.2 Anticipated Contribution

The following serves to present key contributions of this research effort to the transportation community.

- **Developed a methodology for integrating real-time detector streams into a microscopic simulation model.** These detector streams appropriately populated the microscopic simulation (VISSIM) model, according to the vehicle information

contained in the streams. The method's robustness facilitates the integration of data streams from a variety of detector types. Throughout this research detectors types ranged from PC based detectors to VDS (video detection systems).

- **Demonstrated the feasibility of implementing a real-time simulation framework.** This research indentified and capitalized on the opportunities presented by advancements in traffic simulation packages, wireless communication, and detector technologies to implement and evaluate a real-time simulation framework. This was accomplished through successfully creating a series of "real-world" test-beds that were compatible with the real-time data-driven simulation technique that was developed.

- **Illustrated capabilities of real-time simulation.** The results from the experiments, primarily the field experiments, indicated that the proposed methodology is able to accurately estimate field performance measures. Although travel-time, and to a lesser extent – vehicle flow, were the featured performance measures throughout this research, it is anticipated that other measures will be estimated with a similar level of accuracy.

- **Contributed to the understanding of field data needs for real-time simulation.** Successfully implementing a real-time simulation methodology is highly dependent on the availability and accuracy of field data. These data include vehicular volumes, detailed signal timing information, and turning movement distribution. Although static, historical data may be relatively sufficient for signal timing and turning movement distributions, real-time transmission of individual vehicle records is important to the viability of the proposed method. As was demonstrated in the pseudo field experiment, the accuracy in travel estimates was increased when real-time turning

movements and signal timing information were implemented in a real-time fashion. In addition, this research brought to the forefront the importance of signal synchronization between the field signals and those in the simulated environment

- **Informed future real-time simulation efforts of issues and limitations that may be associated with such an implementation.** These included difficulties associated with appropriately accounting for pedestrians (and their behavior), pedestrian vehicle interactions, permissive vehicle movements, and nuanced roadway features (such as grade), that affect traffic performance. These difficulties also spanned the ability to collect and correctly input field data in to the simulation environment to facilitate accurately modeling the study area. These data included signal timing plans, time-of-day turning movement distributions, and an appropriate estimate of desired drive speeds.

- **Developed an advanced simulation calibration procedure and underscored the importance for proper calibration in order to obtain accurate estimates of field travel time.** This procedure is based on a two-part criteria, where the first ensures that simulated flow measures are within a reasonable range of field measures, and the second statistically (and rigorously) evaluates the distribution (and median) of travel time measures.

- **Demonstrated the need for real-time / time-based calibration algorithm, when simulating in real-time or for more than a single time period.** In applying the developed advanced calibration procedure, it was found that different sets of calibration parameter values produced different calibrated models, for different time periods. A real-time algorithm will advance the current calibration practice to models calibrated for current field conditions.

- **Provided additional insights and recommendations regarding parameters to calibrate and the selection of values for these parameters in order to promote the development of a calibrated, arterial simulation model.** There were four parameters that this research recommends modelers begin with in order to calibrate a VISSIM (arterial) model. These parameters were 1) average standstill distance (*ax*), 2) the additive part of the safety distance (*bx_add*), 3) the multiplicative part of safety distance (*bx_mult*), and 4) desire speed range (*dsd_rng*). In terms of the values for these parameters, the values for *ax*, *bx_add*, and *bx_mult* should be chosen so that the value of the average distance between vehicle (d) should be between 26% and 53% of its maximum value. As for the value of *dsd_rng*, 10 mph is recommended. However, the success of such a value is dependent on the accuracy of the field estimate of desired speed. The following sections serve to present the key contributions of this research effort.

APPENDIX A.

BOOTSTRAPPED HEADWAY DATA

Table 24 Bootstrapped Field Startup Flow Data and Five Resamples

Measure	Startup Flow	Resample #1	Resample #2	Resample #3	Resample #4	Resample #5
1	1499	1338	1518	1396	1465	1386
2	1273	1723	1338	1245	1273	1467
3	1467	1286	1723	1273	1402	1465
4	1442	1411	1525	1525	1499	1286
5	1396	1336	1411	1698	1453	1033
6	1518	1402	1336	1592	1100	1245
7	1479	1338	1592	959	1273	1033
8	1456	1456	1338	1430	1386	1336
9	1472	1499	1430	1592	1453	1386
10	1732	1396	1100	1296	1456	1100
11	1411	1296	1430	1698	1273	1100
12	1698	1472	1296	959	1592	1472
13	1336	1442	1402	1479	1472	1592
14	1592	1723	1456	1723	1396	1100
15	1386	1033	1499	1499	1479	1033
16	959	1472	1499	1273	1442	1273
17	1430	1273	1525	1499	1245	1479
18	1465	1245	1732	1296	959	1456
19	1338	1411	1467	1698	1296	1430
20	1402	1456	1386	1525	1698	1430
21	1723	1273	1336	1467	1336	1525
22	1033	1411	1286	1273	1518	1732
23	1245	1698	1286	1396	1698	1698
24	1286	1033	1453	1453	1698	959
25	1100	1499	1033	1467	1442	1453
26	1525	1698	1411	1499	1336	959
27	1296	1525	1338	1723	1499	1430
28	1453	1465	959	1465	1245	1499
Mean	1408	1415	1397	1443	1407	1334
SD	182	172	173	199	168	219

Table 25 Bootstrapped Field Saturation Flow Data and Five Resamples

Measure	Saturation Flow	Resample #1	Resample #2	Resample #3	Resample #4	Resample #5
1	1883	1640	1652	1613	1304	1511
2	1694	1762	1652	1807	1881	1578
3	1229	1613	1578	1652	2480	1652
4	1807	1613	1229	2480	1647	2480
5	1652	1824	1229	2480	1229	1640
6	1361	1807	1807	1613	1652	1229
7	1881	1652	1694	1583	1881	1640
8	1578	1511	2480	1229	1361	1807
9	1583	1652	1694	1361	1583	1881
10	1511	1883	1652	1578	1807	1613
11	1304	1361	1762	1304	1647	1613
12	1640	1647	2480	1640	1881	1652
13	2480	1578	1578	1361	2480	1824
14	1824	1881	1361	1511	1613	1694
15	1613	2480	2480	1824	1304	1640
16	1762	1361	2480	1652	1511	1304
17	1647	1647	1694	1881	1807	1578
Mean	1673	1701	1794	1681	1710	1667
SD	281	252	425	351	358	265

APPENDIX B.

CONSTRUCTING BOOTSTRAP CONFIDENCE INTERVALS

The following will detail the four methods that were used to construct the bootstrap confidence intervals in Chapter 5. These methods include the normal approximation, percentile, bias-corrected and percentile-*t* method. The details associated with each of these methods is adapted from Mooney and Duval (1993) [54]. A sample of the data that was used to create each of these confidence intervals is shown in Appendix A. Below are a few symbols, and their meanings, that will be used to describe how each method produced its confidence interval:

B	number of times data are re-sampled; B= 10000
$\overline{X}$	sample mean
$\tilde{\theta}$	sample statistic; average flow
$\tilde{\sigma}_{\tilde{\theta}}$	standard error (deviation) in sample statistic
$\tilde{\theta}^*$	bootstrap statistic; average flow, from bootstrap data
$\sigma^*_{\tilde{\theta}}$	bootstrapped standard error in bootstrap statistic (average flow)
$\tilde{\theta}^*_b$	bootstrapped data point; flow estimate

Each of these methods were implemented using the R Project for Statistical Computing [56]. The R-Script that was used to construct each of the associated confidence intervals is presented at the end of this appendix.

Normal Approximation Method

The construction of a bootstrap confidence interval using the normal approximation method is similar to the parametric method of creating a confidence interval (Equation A.1). This method is particularly applicable when the statistic being estimated is normally distributed. This method is therefore suitable to estimate confidence intervals for this situation. According to the Lilliefors normality test the startup and saturation flow measurements are hypothesized to be normally distributed, and as such the suitability of the method. However, to maintain the desired robustness throughout the calibration process, flow measures are not required to be normally distributed, and may therefore render this method unsuitable. The normal approximation method is included as a part of this effort as an exploration exercise in order to sufficiently compare confidence intervals from all four methods.

Equation A.2 demonstrates the parallel between itself and the parametric method of forming a confidence interval; and how one calculates the confidence interval, using the normal approximation method

$$CI = \bar{X} \pm z_{\alpha/2}\sigma_{\bar{X}} \qquad \textbf{Equation A.1}$$

$$CI_{boot} = \tilde{\theta} \pm z_{\alpha/2}\tilde{\sigma}^*_{\tilde{\theta}} \qquad \textbf{Equation A.2}$$

where: $\tilde{\sigma}^*_{\tilde{\theta}} = \left[\left(\sum\left(\tilde{\theta}^*_b - \tilde{\theta}^*_{(.)}\right)^2\right)/(B-1)\right]^{1/2}$ and $\tilde{\theta}^*_{(.)} = \sum \tilde{\theta}^*_b / B$

In applying the above equations, the following presents the construction of a 95% confidence interval for startup flow:

$$CI_{boot_{2-5}} = 1408 \pm 1.96 \times \tilde{\sigma}^*_{\tilde{\theta}}$$

$$\tilde{\sigma}_{\tilde{\theta}}^{*} = \left[\left(\sum\left(\tilde{\theta}_{b}^{*} - \tilde{\theta}_{(.)}^{*}\right)^{2}\right)/(10000-1)\right]^{1/2} = 33.9$$

$$CI_{boot_{2-5}} = 1408 \pm 1.96 \times 33.9 = (\mathbf{1341}, \mathbf{1474})$$

Following the same procedure as above, the 95% confidence interval for saturation flow:

$$CI_{boot_{6-9}} = 1673 \pm 1.96 \times \tilde{\sigma}_{\tilde{\theta}}^{*}$$

$$\tilde{\sigma}_{\tilde{\theta}}^{*} = \left[\left(\sum\left(\tilde{\theta}_{b}^{*} - \tilde{\theta}_{(.)}^{*}\right)^{2}\right)/(10000-1)\right]^{1/2} = 66.4$$

$$CI_{boot_{6-9}} = 1673 \pm 1.96 \times 66.4 = (\mathbf{1544}, \mathbf{1805})$$

The Percentile Method

The percentile method is the simplest of the four methods. It assumes that a bootstrap statistic is an unbiased estimate of the same statistic for the sample data, and that the sample's CDF is approximately that of the bootstrapped data. Therefore, to form a 95% confidence interval, the bootstrapped data was first ordered and the data point associated with the 2.5th and the 97.5th percentile represent the lower and upper bounds of the confidence interval, respectively. In this case, after ordering the 10,000 flow estimates, the 250th and the 9750th values, make up the end points of the 95% confidence interval.

Applying the percentile method:

$$CI_{boot_{2-5}} = (boot\_data\_2\_5[250], bootdata\_2\_5[9750]) = (\mathbf{1340}, \mathbf{1473})$$

$$CI_{boot_{6-9}} = (boot\_data\_6\_9[250], bootdata\_6\_9[9750]) = (\mathbf{1552}, \mathbf{1811})$$

In addition to being the most direct method, another advantage is that the percentile method's confidence interval is also capable successfully forming an interval around a statistic that is asymmetrically distributed – unlike the normal approximation method. A disadvantage of this method is that poorly formed intervals may be obtained for small datasets.

The Bias-Corrected Method

The Bias Corrected (BC) method eliminates the assumption that the bootstrap statistic is an unbiased estimate of the sample statistic, and that the sample statistic is an unbiased estimate of the population statistic. Instead, this method assumes there is a monotonic transformation that involves a biasing constant, z_o. This constant is the adjustment to the bootstrapped statistic that will facilitate the appropriate representation of the sample statistic. The construction of a confidence interval, using the BC method, involves two steps: 1) calculate z_o and 2) use z_o to adjust the bootstrap distribution. See below.

Step 1: Calculate z_o :

$$z_0 = \Phi^{-1}\{pr(\tilde{\theta}^* \leq \tilde{\theta})\}$$

where Φ is the cumulative distribution function for the standard normal variable.

Step 2: Use z_o to adjust the bootstrap distribution, i.e. adjust the percentile values of $\tilde{\theta}^*$ by the use of z_o .

Lower Bound of BC confidence interval = the value of the $\tilde{\theta}^*$ at the

$$= \left[\{\Phi(2z_0 - z_{\alpha/2})\} \times 100\right]\text{percentile}$$

Upper Bound of BC confidence interval = the value of the $\tilde{\theta}^*$ at the

$$= \left[\{\Phi(2z_0 + z_{\alpha/2})\} \times 100\right]\text{percentile}$$

Applying the above steps to the startup flow measures:

$$z_0 = \Phi^{-1}\{pr(\tilde{\theta}^* \leq 1408)\} = \Phi^{-1}\{0.51\} = 0.02$$

Percentile for Lower Bound of BC confidence interval =

$$= \{\Phi(2(0.02) - 1.96)\} \times 100 = \Phi(1.92) \times 100 = 2.7$$

2.7th percentile of $\tilde{\theta}^*$=1340

Percentile for Upper Bound of BC confidence interval =

$$= \{\Phi(2(0.02) + 1.96)\} \times 100 = \Phi(2) \times 100 = 97.7$$

97.7th percentile of $\tilde{\theta}^*$=1473

$$CI_{boot_{2-5}} = (\mathbf{1340}, \mathbf{1473})$$

Applying the above steps to the saturation flow measures:

$$z_0 = \Phi^{-1}\{pr(\tilde{\theta}^* \leq 1673)\} = \Phi^{-1}\{0.51\} = 0.02$$

Percentile for Lower Bound of BC confidence interval =

$$= \{\Phi(2(0.02) - 1.96)\} \times 100 = \Phi(1.92) \times 100 = 2.7$$

2.7th percentile of $\tilde{\theta}^*$=1552

Percentile for Upper Bound of BC confidence interval =

$$= \{\Phi(2(0.02) + 1.96)\} \times 100 = \Phi(2) \times 100 = 97.7$$

97.7th percentile of $\tilde{\theta}^*$=1811

$$CI_{boot_{6-9}} = (\mathbf{1552}, \mathbf{1811})$$

Although the BC method is advantageous in that it eliminates the unbiased assumption, there are still disadvantages of this method. One of the more significant disadvantage is that the construction of the confidence interval its centered on a value that

may not be at all pivotal for the population being studied. This is particularly a problem in the event of skewed sample data.

The Percentile-*t* Method

The percentile-*t* method produces the most accurate confidence intervals. It is however, the most computationally expensive method as it involves double bootstrapping the data. The percentile-*t* method's confidence interval is defined as:

$$p\left(\tilde{\theta} - t^*_{\alpha/2}\tilde{\sigma}_{\tilde{\theta}} < \theta < \tilde{\theta} - t^*_{1-\alpha/2}\tilde{\sigma}_{\tilde{\theta}}\right) = 1 - \alpha$$

where $\tilde{\sigma}_{\tilde{\theta}}$ is calculated for the original sample.

However before one can apply the above formula one has the first determine the values of $t^*_{\alpha/2}$ and $t^*_{1-\alpha/2}$. To do this $\tilde{\theta}^*$ will be transformed to standardized variable *t** by using

$$t^*_b = (\tilde{\theta}^*_b - \tilde{\theta})/\tilde{\sigma}_{\tilde{\theta}}$$

For this transformation, each $\tilde{\theta}^*_b$ will be weighted by an estimate of its standard error. To develop an estimate of the associated standard error, a second bootstrap is applied to the data points that formed each $\tilde{\theta}^*_b$. This second bootstrap effort involved re-sampling the corresponding headway measurements 200 times and calculating the standard deviation. The distribution of t^*_b was then used to supply values for $t^*_{\alpha/2}$ and $t^*_{1-\alpha/2}$ in order to construct this method's confidence interval. The distribution of t^*_b for both the startup and saturation flow measures are shown in Figure 80 and Figure 81.

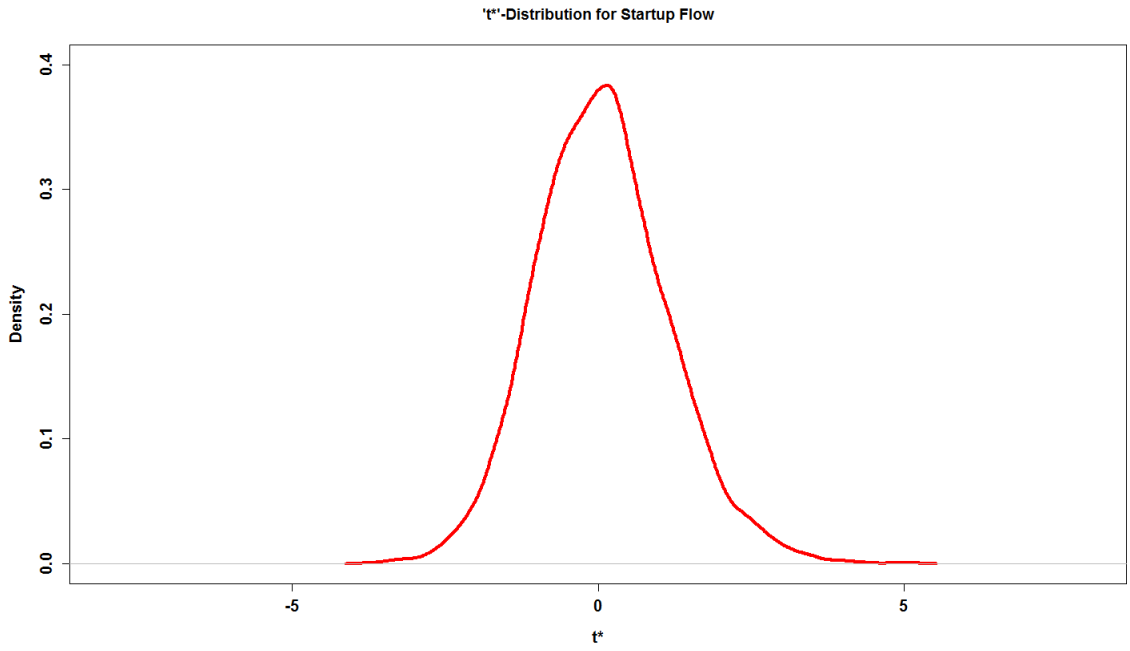


Figure 80 *t-Distribution for Startup Flow**

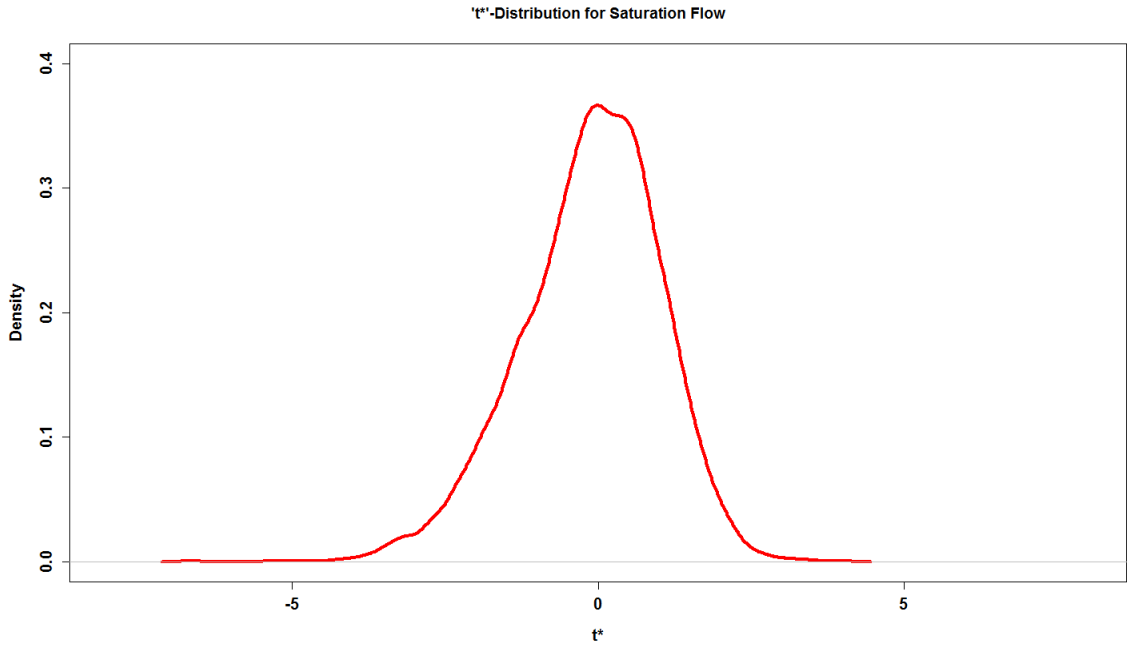


Figure 81 *t-Distribution for Saturation Flow**

Using the above *t** distributions, the confidence intervals produced by the percentile-*t* method are as follows:

$$CI = (\tilde{\theta} + t^*_{\alpha/2}\tilde{\sigma}_{\tilde{\theta}}, \tilde{\theta} + t^*_{1-\alpha/2}\tilde{\sigma}_{\tilde{\theta}})$$

$$CI_{boot_{2-5}} = (1408 + (-1.93) \times 33.9, 1408 + (2.36) \times 33.9) = (\mathbf{1342}, \mathbf{1488})$$

$$CI_{boot_{6-9}} = (1673 + (-2.63) \times 66.4, 1673 + (1.84) \times 66.4) = (\mathbf{1499}, \mathbf{1796})$$

Although the percentile-*t* method does provide the most accurate confidence interval, the analyst should be sure to carefully evaluate the purpose for which the interval was created and then decide its appropriateness.

R-Script Method Implementation

Below is the R-Script that was used to implement the above methods to construct the various confidence intervals.

```
sat_flow_ngsim_2_5=c(1499,
1273,1467,1442,1396,1518,1479,1456,1472,1732,1411,1698,1336,1592,1386,
959, 1430,1465,1338,1402,1723,1033,1245,1286,1100,1525,1296,1453)
sat_flow_ngsim_6_9=c(1883,
1694,1229,1807,1652,1361,1881,1578,1583,1511,1304,1640,2480,1824,
1613,1762,1647)

mn_sat_flow_ngsim_2_5=mean(sat_flow_ngsim_2_5)
mn_sat_flow_ngsim_6_9=mean(sat_flow_ngsim_6_9)
sd_sat_flow_ngsim_2_5=sd(sat_flow_ngsim_2_5)
sd_sat_flow_ngsim_6_9=sd(sat_flow_ngsim_6_9)
sf<-read.csv("Sat Flows - Bootstrap.txt",header=T, nrow=10000)

B=length(sf$X2)

###normal approximation method ###
CI_NA_2_5_1_new=mean(bt_mean_2_5)-1.96*sd(bt_mean_2_5)
CI_NA_2_5_2_new=mean(bt_mean_2_5)+1.96*sd(bt_mean_2_5)
CI_NA_6_9_1_new=mean(bt_mean_6_9)-1.96*sd(bt_mean_6_9)
CI_NA_6_9_2_new=mean(bt_mean_6_9)+1.96*sd(bt_mean_6_9)
### percentile method ###
CI_PM_2_5_1_new=quantile(bt_mean_2_5,seq(0,1,0.005))[6]
CI_PM_2_5_2_new=quantile(bt_mean_2_5,seq(0,1,0.005))[196]
CI_PM_6_9_1_new=quantile(bt_mean_6_9,seq(0,1,0.005))[6]
CI_PM_6_9_2_new=quantile(bt_mean_6_9,seq(0,1,0.005))[196]
### Bias Correctd Method for CI ###
```

```
z_score_tbl=as.matrix(read.csv("Standard Normal Probability Table.txt",
header=F,skip=1)[1:35,2:
11])
prob_new=(length(which(quantile(bt_mean_2_5,seq(0,1,0.005))<=mn_sat_flow
_ngsim_2_5))-1)/200
if(prob_new<0.5)
{prob_new=1-prob_new}
new_z_new=(z[which.min(abs(z_prop-prob_new))]*2 + 1.96)
per_cent_new=z_prop[which.min(abs(z-new_z_new))]
CI_BC_2_5_2_new=quantile(bt_mean_2_5,seq(0,1,0.005))[ceiling(per_cent_ne
w/.005)]
### CI_BC_2_5_2 = Upper Endpoint , CI_BC_2_5_1 = Lower Endpoint
new_z_2_new=abs(z[which.min(abs(z_prop-prob_new))]*2 - 1.96)
per_cent_2_new=1-z_prop[which.min(abs(z-new_z_2_new))]
CI_BC_2_5_1_new=quantile(bt_mean_2_5,seq(0,1,0.005))[ceiling(per_cent_2_
new/.005)]
prob_new=(length(which(quantile(bt_mean_6_9,seq(0,1,0.005))<=mn_sat_flow
_ngsim_6_9))-1)/200
if(prob_new<0.5)
{prob_new=1-prob_new}
new_z_new=(z[which.min(abs(z_prop-prob_new))]*2 + 1.96)
per_cent_new=z_prop[which.min(abs(z-new_z_new))]
CI_BC_6_9_2_new=quantile(bt_mean_6_9,seq(0,1,0.005))[ceiling(per_cent_ne
w/.005)]
### CI_BC_6_9_2 = Upper Endpoint , CI_BC_2_5_1 = Lower Endpoint
new_z_2_new=abs(z[which.min(abs(z_prop-prob_new))]*2 - 1.96)
per_cent_2_new=1-z_prop[which.min(abs(z-new_z_2_new))]
CI_BC_6_9_1_new=quantile(bt_mean_6_9,seq(0,1,0.005))[ceiling(per_cent_2_
new/.005)]

### t-Percentile Methond for CI ###
bt_mean_2_5=bt_sd_2_5=bt_bt_sd_2_5=0
for (i in 1:10000)
{
      bt_mean_2_5_j=bt_bt_mean_2_5_k=bt_bt_mean_2_5_l=0
      for(j in 1 : length(sat_flow_ngsim_2_5))
      {
            bt_mean_2_5_j[j] =
            sat_flow_ngsim_2_5[sample(1:length(sat_flow_ngsim_2_5),1)]
      }
      for(k in 1:200)
      {
            for (l in 1:length(bt_mean_2_5_j))
            {
                  bt_bt_mean_2_5_l[l]=bt_mean_2_5_j[sample(1:length(bt_
                  mean_2_5_j),1)]
            }
            bt_bt_mean_2_5_k[k]=mean(bt_bt_mean_2_5_l)
      }
      bt_bt_sd_2_5[i]=sd(bt_bt_mean_2_5_k)
      bt_mean_2_5[i] = mean(bt_mean_2_5_j)
      bt_sd_2_5[i] = sd(bt_mean_2_5_j)
}
t_b_2_5_new=(bt_mean_2_5-mn_sat_flow_ngsim_2_5)/bt_bt_sd_2_5

CI_Pct_t_2_5_1_new=mn_sat_flow_ngsim_2_5+(t_b_2_5_new[order(t_b_2_5_new
      )][250]*sd(bt_mean_2_5))
CI_Pct_t_2_5_2_new=mn_sat_flow_ngsim_2_5+(t_b_2_5_new[order(t_b_2_5_new
      )][9750]*sd(bt_mean_2_5))
bt_mean_6_9=bt_sd_6_9=bt_bt_sd_6_9=0
for (i in 1:10000)
{
      bt_mean_6_9_j=bt_bt_mean_6_9_k=bt_bt_mean_6_9_l=0
      for(j in 1 : length(sat_flow_ngsim_6_9))
```

```
      {
            bt_mean_6_9_j[j] =
            sat_flow_ngsim_6_9[sample(1:length(sat_flow_ngsim_6_9),1)]
      }
      for(k in 1:200)
      {
            for (l in 1:length(bt_mean_6_9_j))
            {
                  bt_bt_mean_6_9_l[l]=bt_mean_6_9_j[sample(1:length(bt_m
                  ean_6_9_j),1)]
            }
            bt_bt_mean_6_9_k[k]=mean(bt_bt_mean_6_9_l)
      }
      bt_bt_sd_6_9[i]=sd(bt_bt_mean_6_9_k)
      bt_mean_6_9[i] = mean(bt_mean_6_9_j)
      bt_sd_6_9[i] = sd(bt_mean_6_9_j)
}
t_b_6_9_new=(bt_mean_6_9-mn_sat_flow_ngsim_6_9)/bt_bt_sd_6_9

CI_Pct_t_6_9_1_new=mn_sat_flow_ngsim_6_9+(t_b_6_9_new[order(t_b_6_9_new
      )][250]*sd(bt_mean_6_9))
CI_Pct_t_6_9_2_new=mn_sat_flow_ngsim_6_9+(t_b_6_9_new[order(t_b_6_9_new
      )][9750]*sd(bt_mean_6_9))

plot(density(t_b_6_9_new),xlab="t*", main="'t*'-Distribution for
      Saturation Flow",
      cex.axis=1.25,cex.lab=1.25,font.axis=2,font.lab=2,font.main=2,lwd=
      4,xlim=c(-8,8),ylim=c(0,
      0.4),col=2)

plot(density(t_b_2_5_new),xlab="t*", main="'t*'-Distribution for
      Startup Flow",
      cex.axis=1.25,cex.lab=1.25,font.axis=2,font.lab=2,font.main=2,lwd=
      4,xlim=c(-8,8),ylim=c(0,
     0.4),col=2)-
```

APPENDIX C.

PLOTS OF REPLICATES SATISFING CALIBRATION CRITERIA

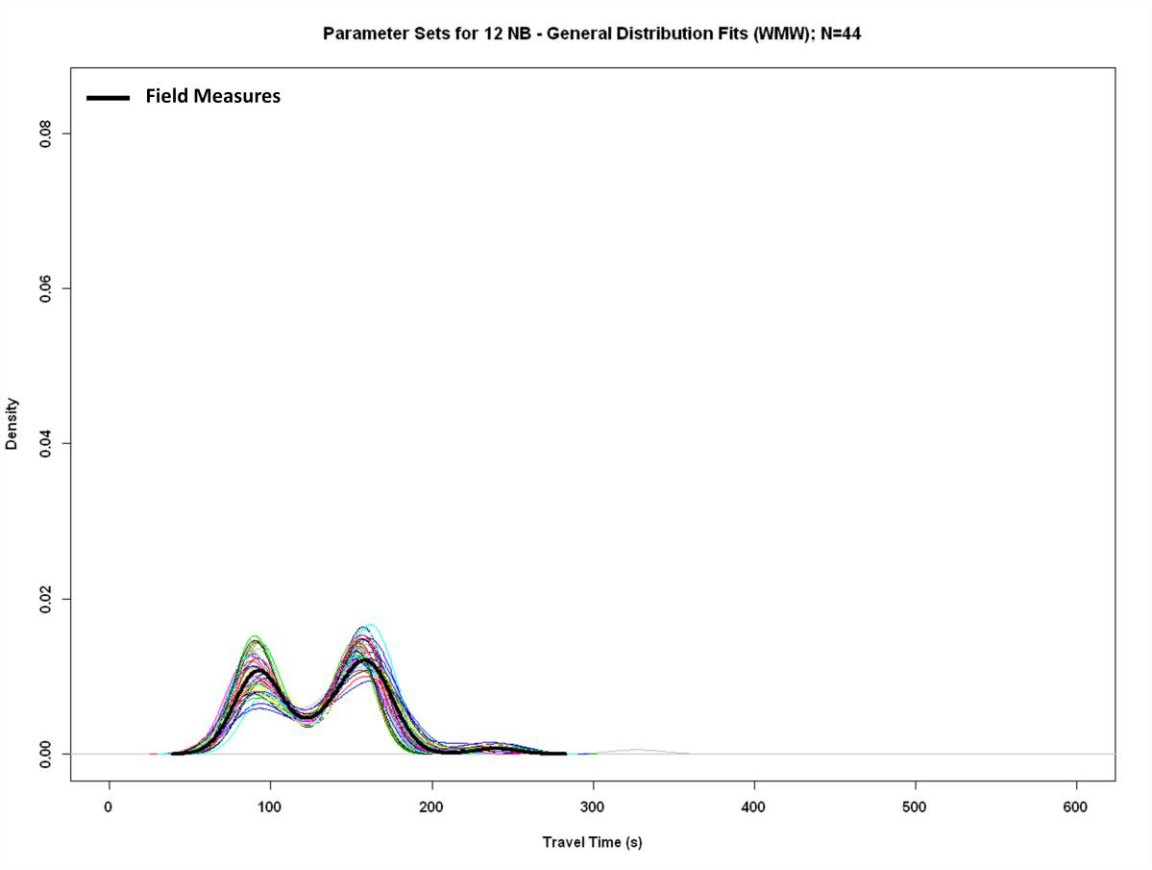


Figure 82 Models from WMW Test, $M_n|U$ – Noon Northbound

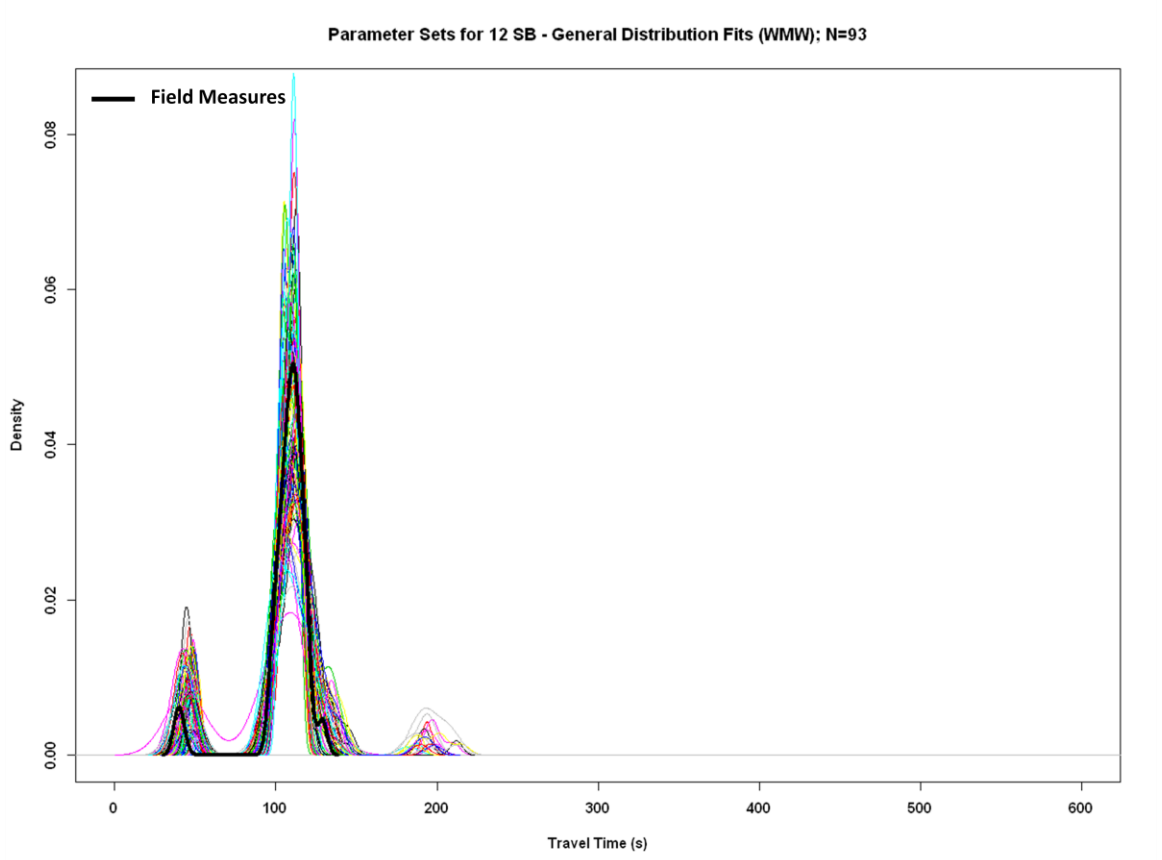


Figure 83 Models from WMW Test, $M_n|U$ – Noon Southbound

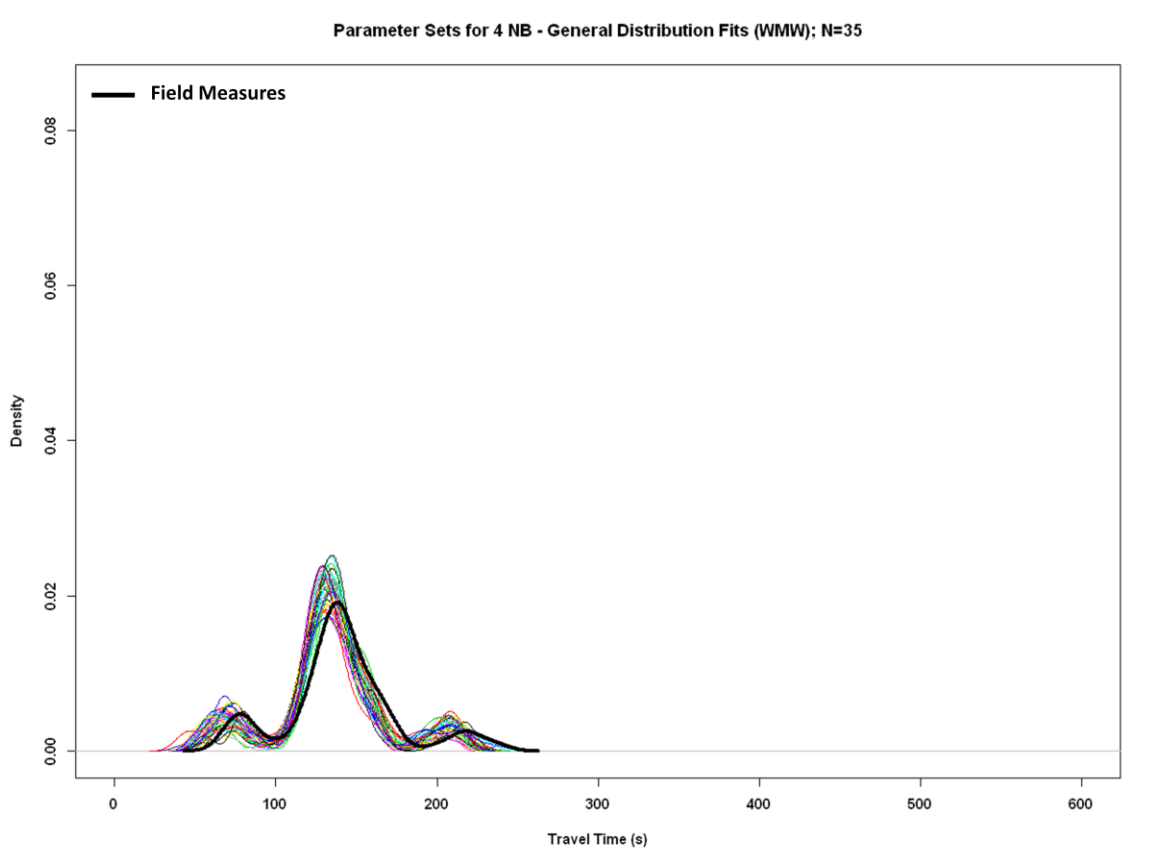


Figure 84 Models from WMW Test, $M_n|U$ – Evening Northbound

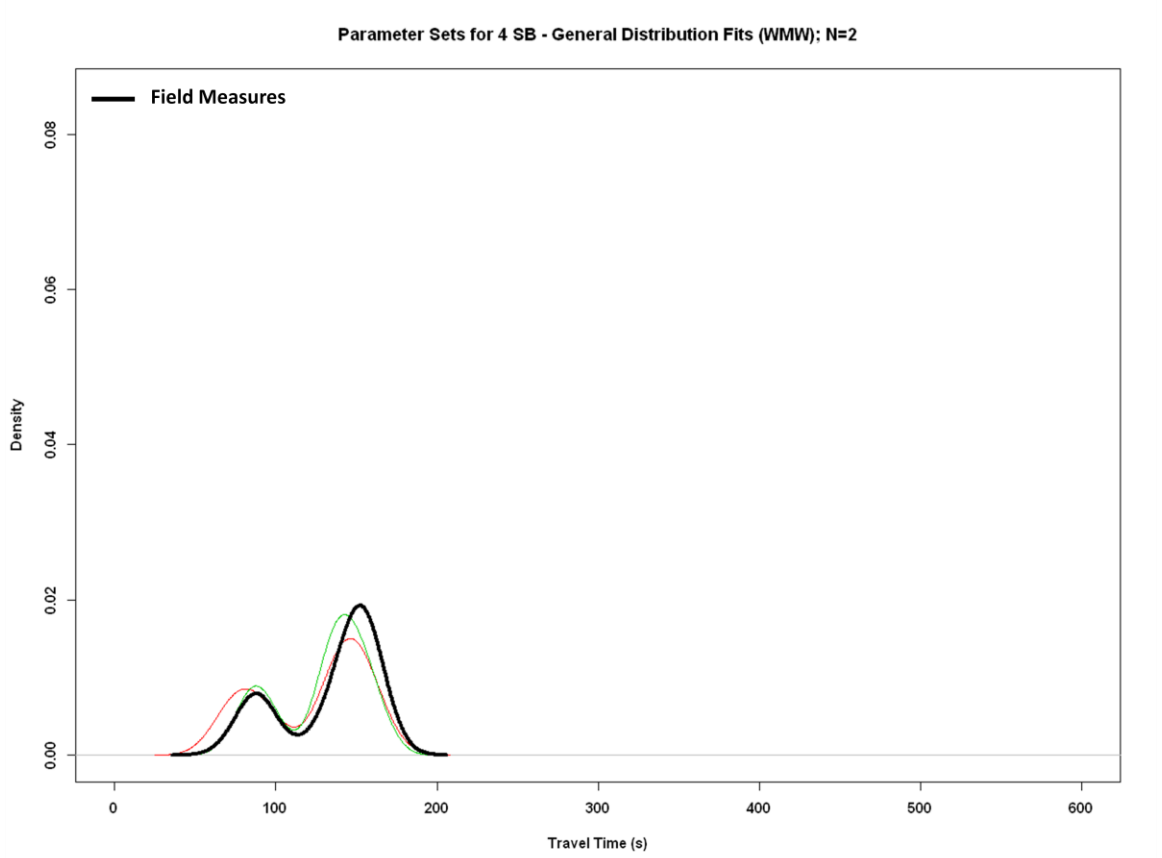


Figure 85 Models from WMW Test, $M_n|U$ – Evening Southbound

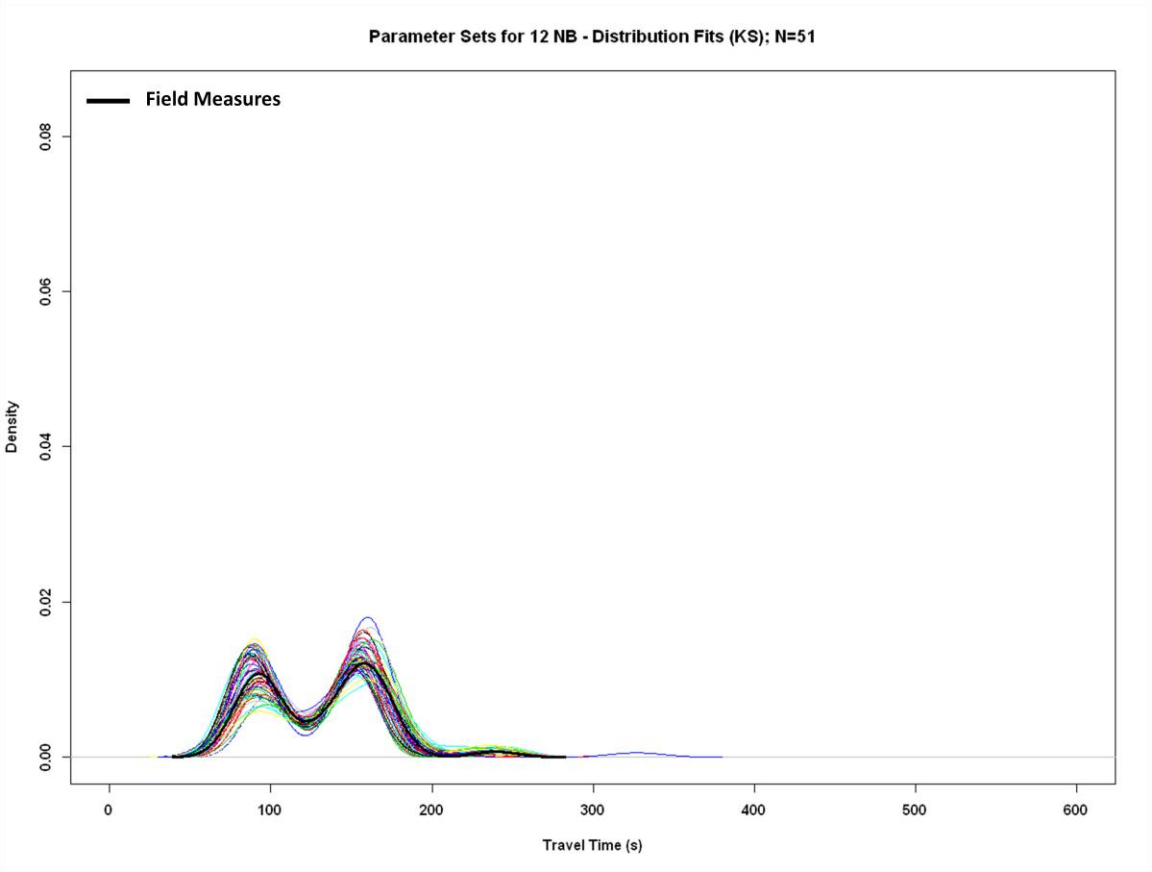


Figure 86 Models from KS Test, $M_n|D$ – Noon Northbound

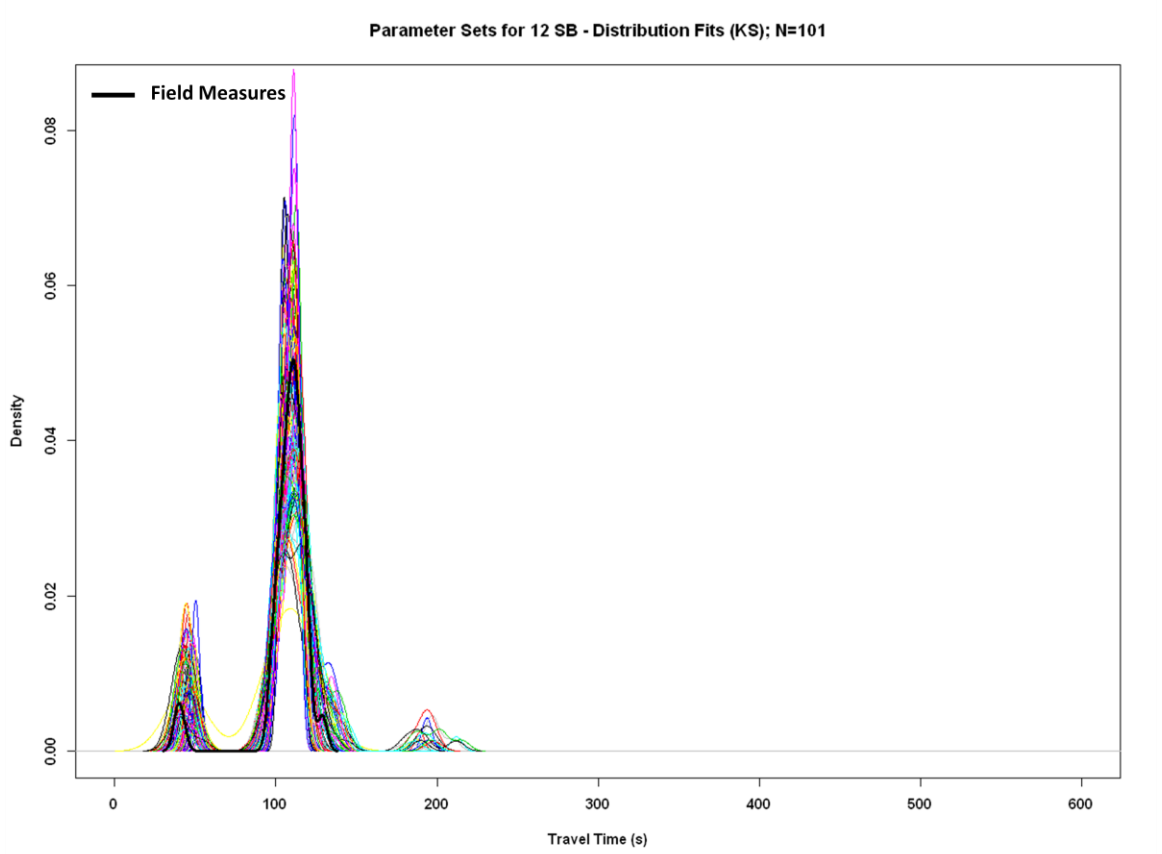


Figure 87 Models from KS Test, $M_n|D$ – Noon Southbound

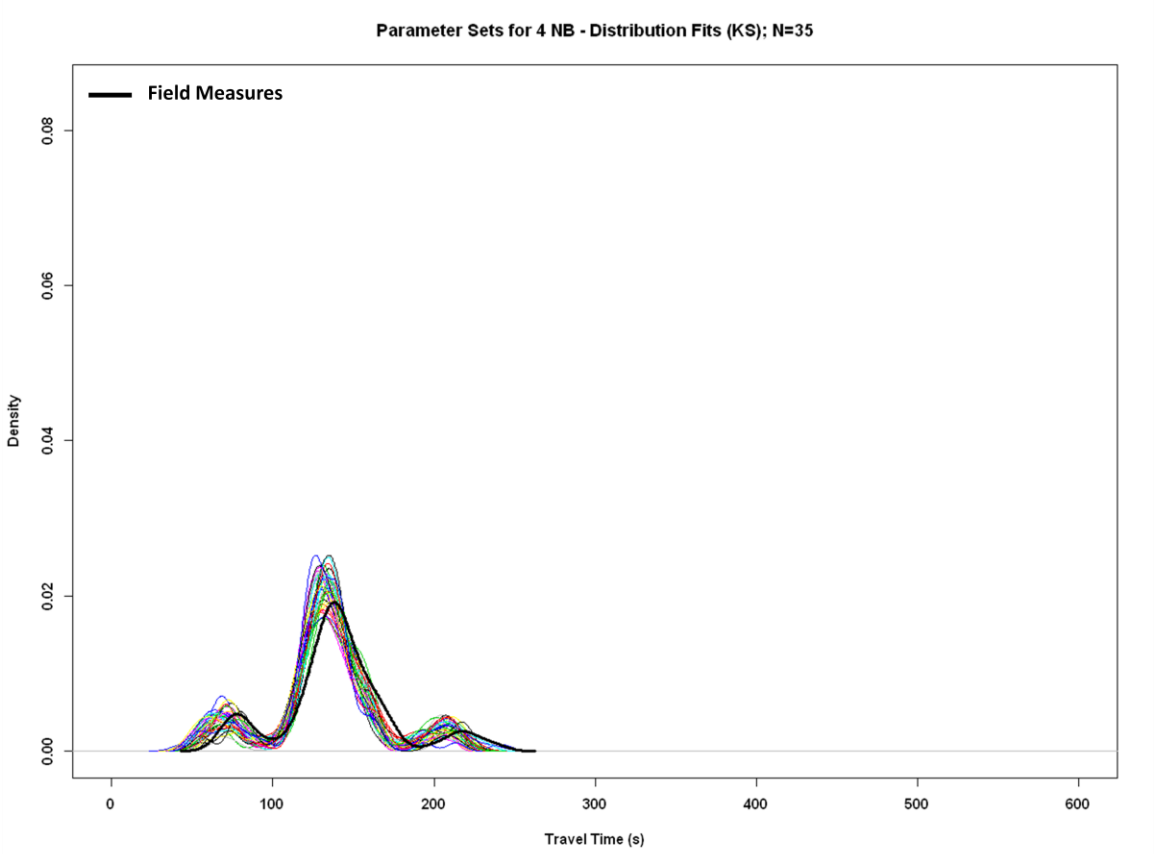


Figure 88 Models from KS Test, $M_n|D$ – Evening Northbound

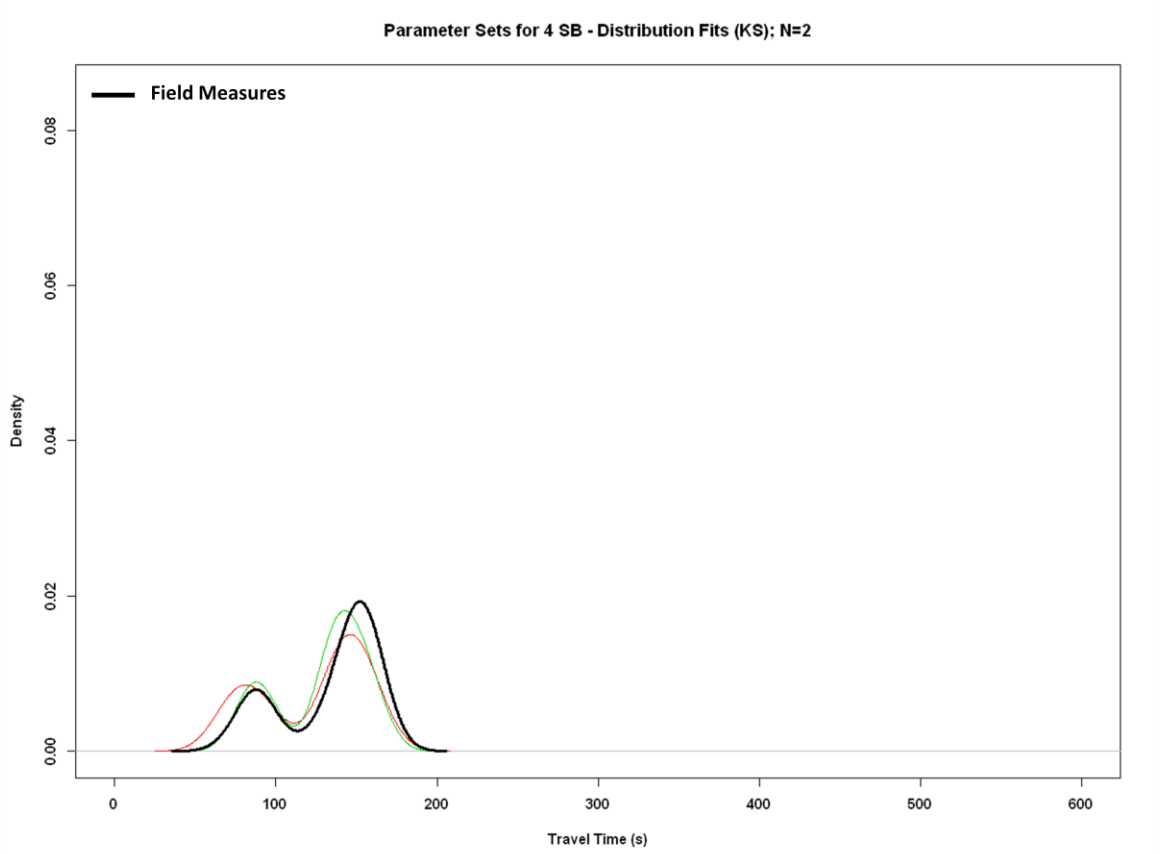


Figure 89 Models from KS Test, $M_n|D$ – Evening Southbound

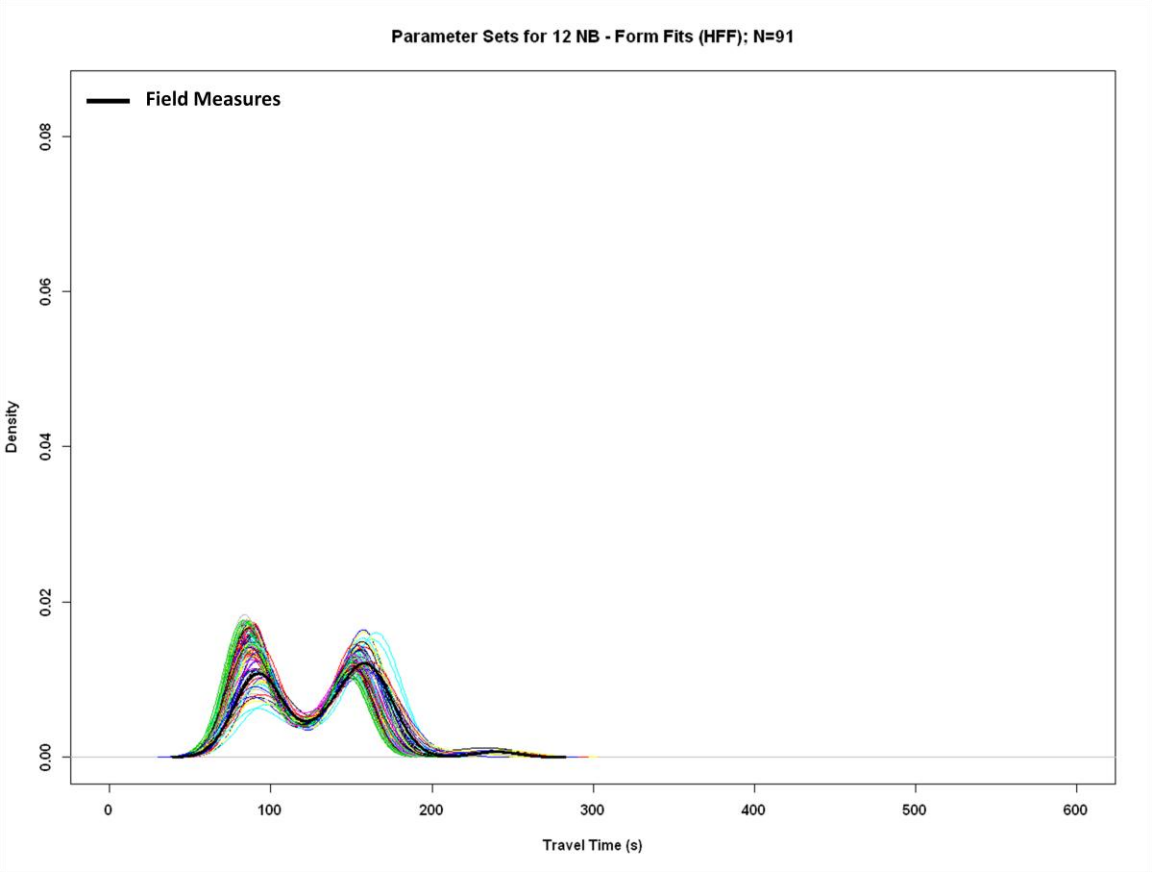


Figure 90 Models from HFF Test, $M_n|H$ – Noon Northbound

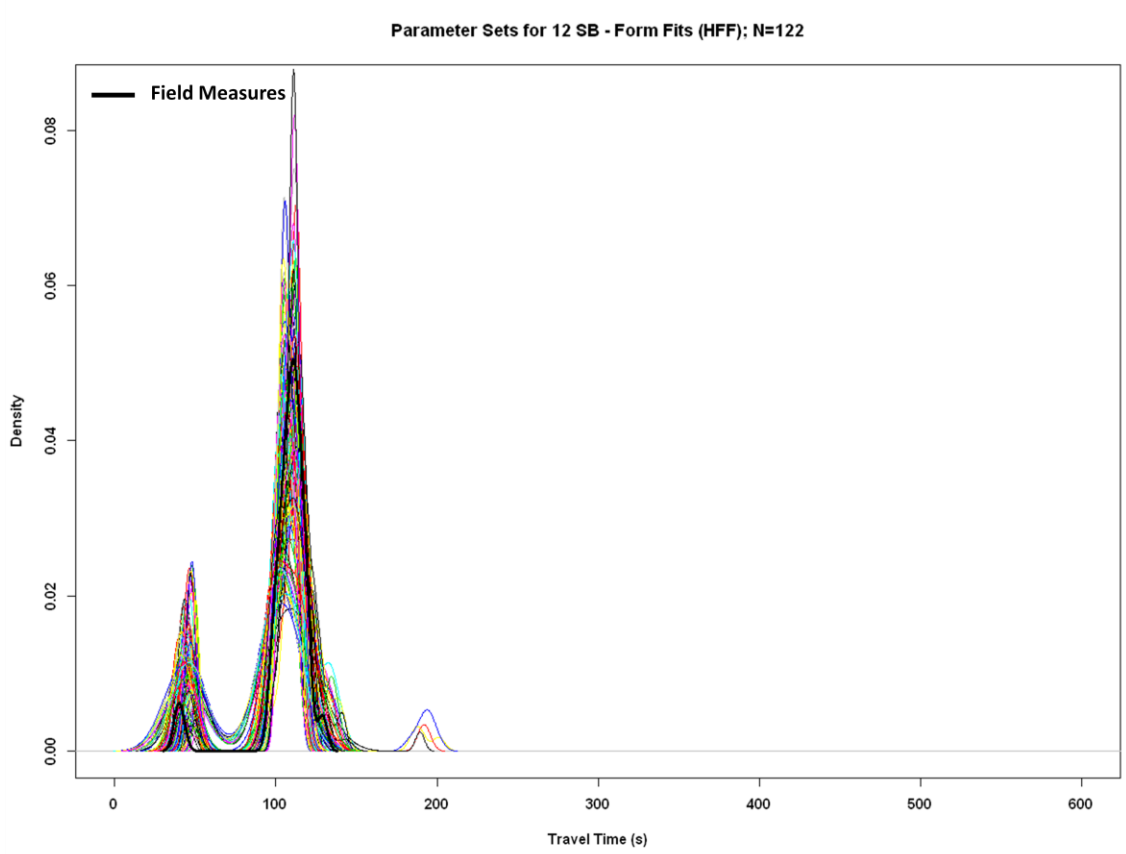


Figure 91 Models from HFF Test, $M_n|H$ – Noon Southbound

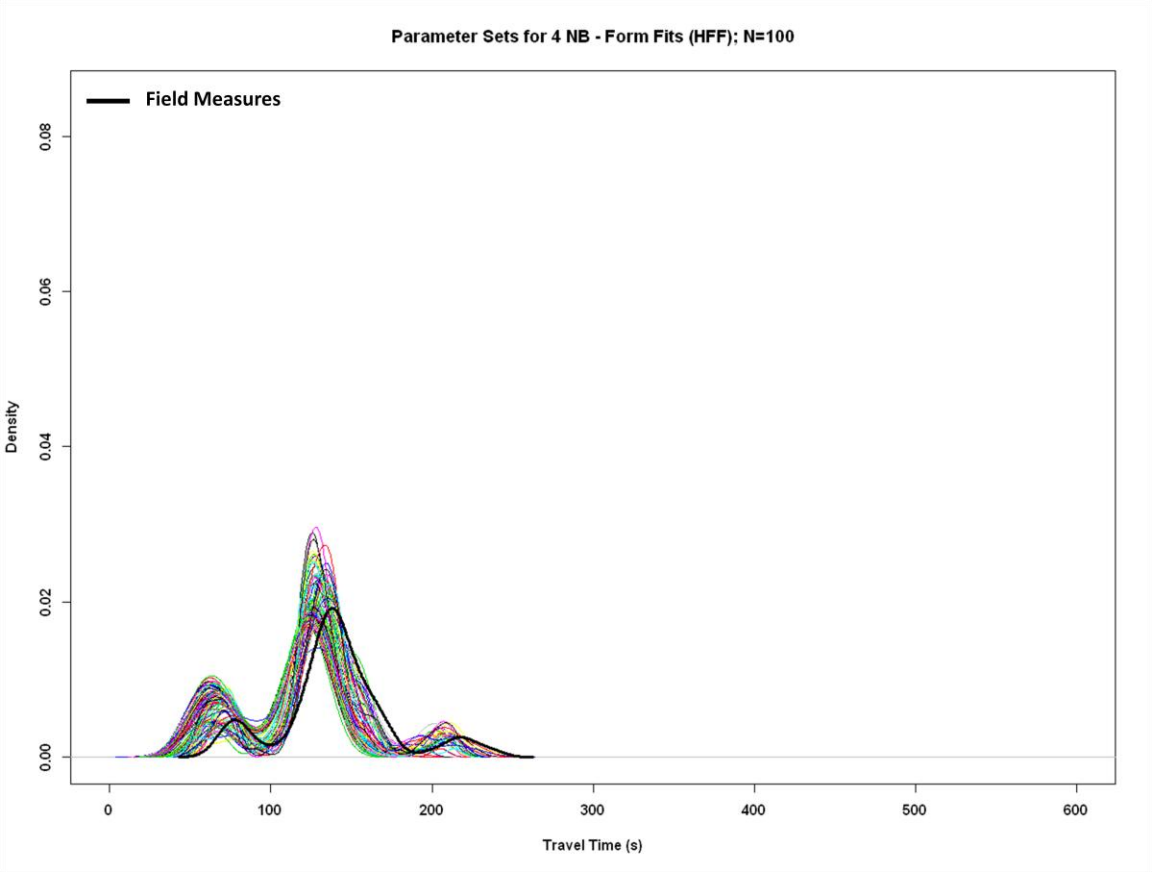


Figure 92 Models from HFF Test, $M_n|H$ – Evening Northbound

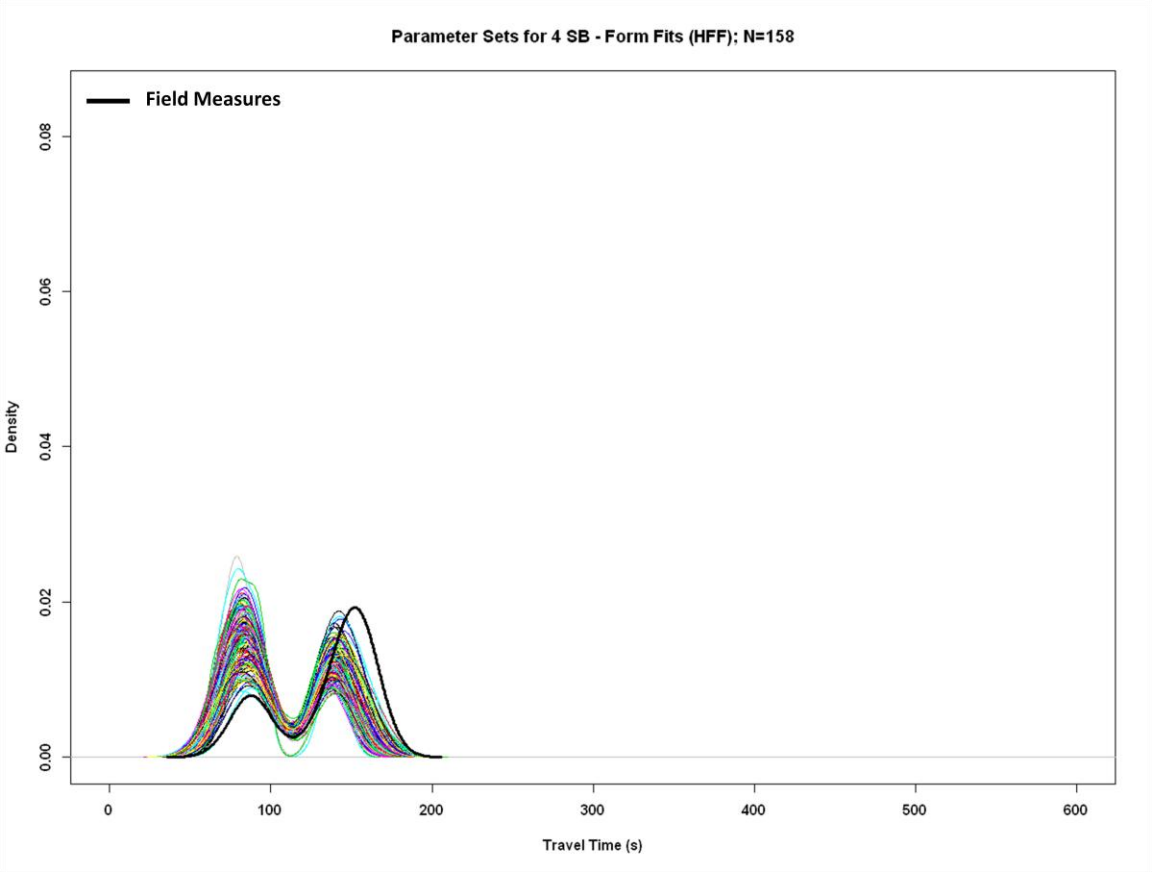


Figure 93 Models from HFF Test, $M_n|H$ – Evening Southbound

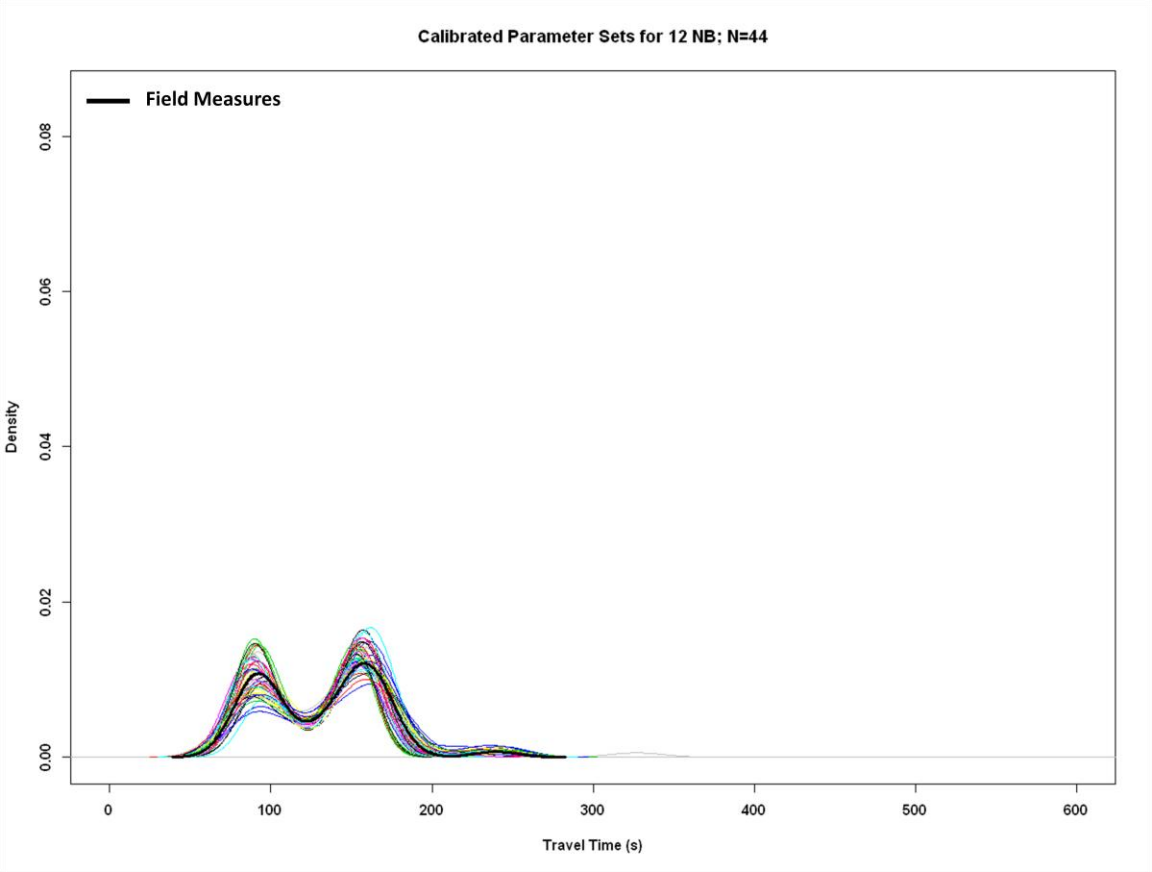


Figure 94 Calibrated Replicates – Noon Northbound

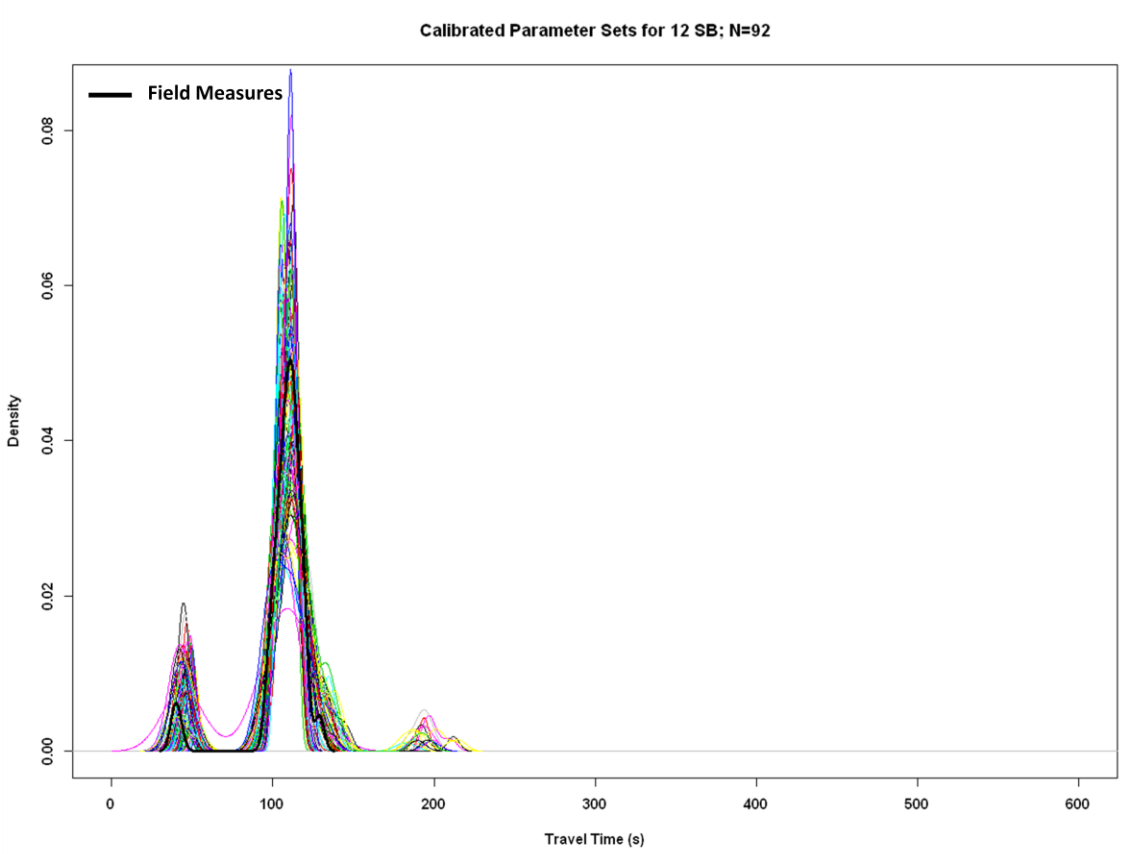


Figure 95 Calibrated Replicates – Noon Southbound

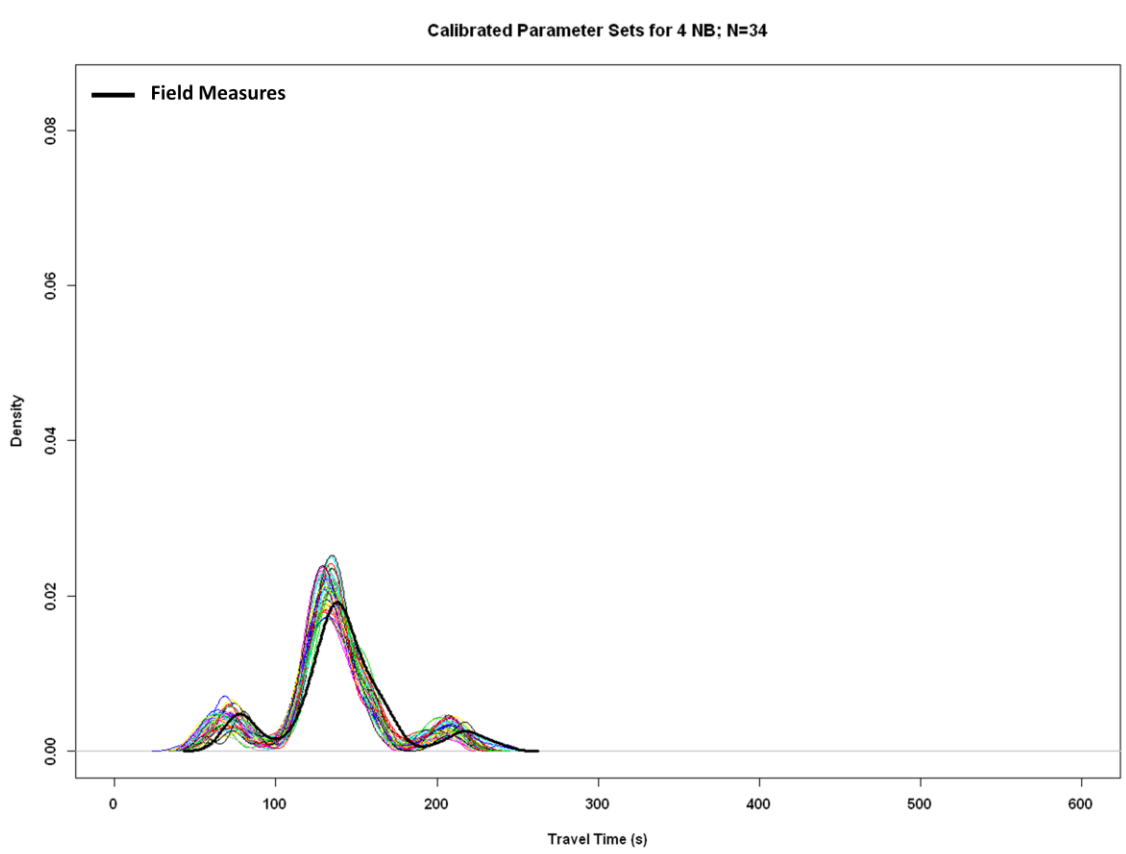


Figure 96 Models Calibrated Replicates – Evening Northbound

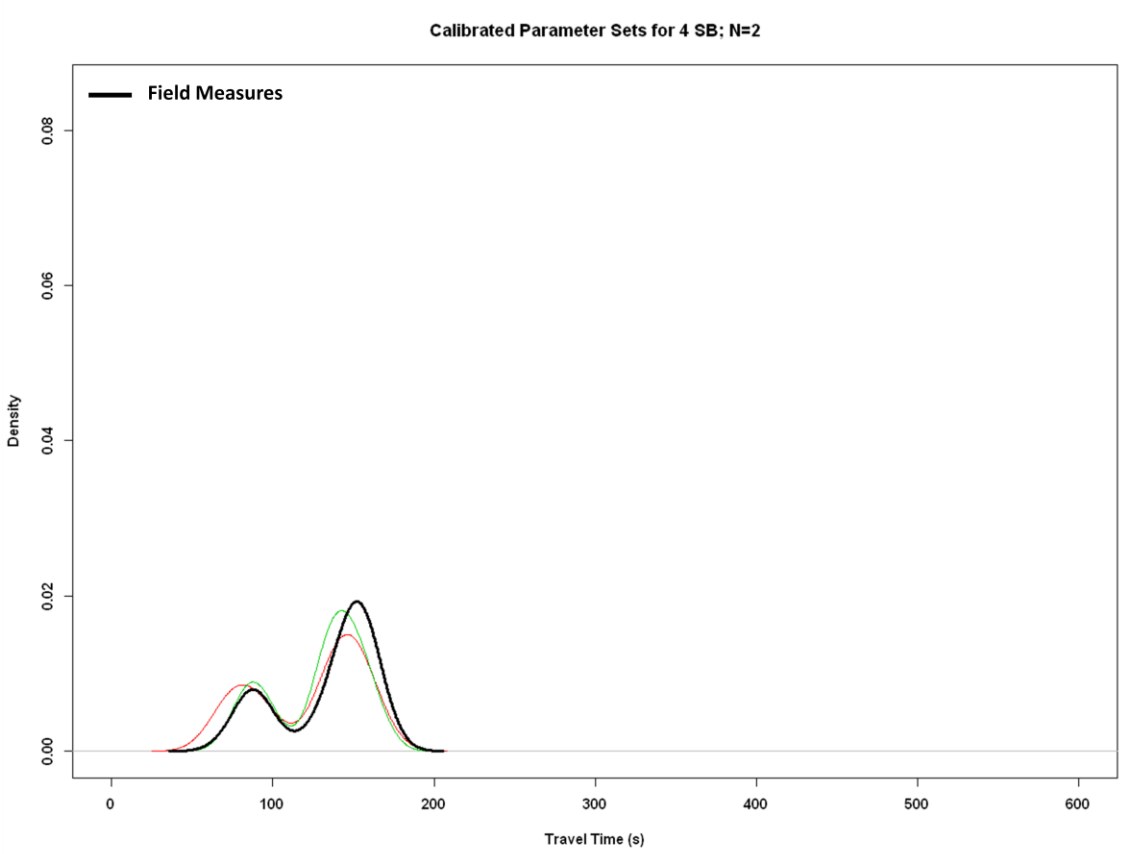


Figure 97 Calibrated Replicates – Evening Southbound

REFERENCES

[1] D. Schrank and T. Lomax, "Urban Mobility Report 2009," *Report for the Texas Transportation Institute*, no. July, 2009.

[2] Research and Innovative Technology Administration, "Intelligent Transportation Systems Benefits, Costs, Deployment, and Lessons Learned 2008 Update," Washinton DC, 2008.

[3] Research and Innovative Technology Administration (RITA), "Intelligent Transportation Systems," Washinton DC, 2009.

[4] Google, "Google Maps," *Web Page*, 2011. [Online]. Available: http://maps.google.com/. [Accessed: 03-Jan-2010].

[5] NAVTEQ, "NAVTEQ Traffic," *Web Page*, 2011. [Online]. Available: http://www.navteq.com/. [Accessed: 03-Jan-2009].

[6] Inrix, "INRIX," 2011. [Online]. Available: http://www.inrix.com/.

[7] P. G. Gipps, "The Estimation of a Measure of Vehicle Delay from Detector Output," Newcastle, 1977.

[8] V. Sisiopiku and N. Rouphail, "Toward the use of Detector Output for Arterial Link Travel Time Estimation: A Literature Review," *Transportation Research Record*, vol. 2, no. 1457, pp. 158-165, 1994.

[9] H. E. Gault and I. G. Taylor, “The Use of Vehicle Detector to Assess Delay in Computer-Controlled Area Traffic Control,” Newcastle, 1977.

[10] H. Zhang and E. Kwon, “Travel Time Estimation on Urban Arterials Using Loop Detector Data,” Iowa City, 1997.

[11] H. Strobel, “Traffic Control Systems Analysis by Means of Dynamic State and Input-Output Models,” Laxenburg, Austria., 1977.

[12] P. Bohnke and E. Pfannerstill, “A system for the automatic surveillance of traffic situations,” *ITE Journal*, vol. 56, no. 1, pp. 41 - 45, 1986.

[13] N. Sensys, “Sensys Networks,” 2012. [Online]. Available: http://www.sensysnetworks.com/. [Accessed: 18-Feb-2012].

[14] T. Usami, K. Ikenoue, and T. Miyasako, “Travel Time Prediction Algorithm and Signal Operations at Critical Intersection Controlling Travel Time,” in *Second International Conference on Road Traffic Control*, p. 1986.

[15] B. O. P. R. BPR, “Traffic Assignment Manual,” Washinton DC, 1964.

[16] H. M. Zhang, “Link-Journey-Speed Model for Arterial Traffic,” *Transportation Research Record*, vol. 1676, no. 1, pp. 109-115, Jan. 1999.

[17] H. X. Liu and X. W. Wenteng Ma, “Time-dependent travel time estimation model for signalized arterial network,” in *86th Transportation Research Board Annual Meeting 2007*, 2007, vol. 1, p. 21.

[18] H. Wang and A. Hobeika, “Travel Time Estimation on Arterial Streets,” Blacksburg, VA, 2007.

[19] D. J. Dailey and F. W. Cathey, “AVL-Equipped vehicles as traffic probe sensors,” Seattle, 2002.

[20] Y. Li and M. McDonald, “Link travel time estimation using single GPS equipped probe vehicle,” in *The IEEE 5th International Conference on Intelligent Transportation Systems, 2002. Proceedings*, 2002, pp. 932–937.

[21] K. Choi and Y. Chung, “A data fusion algorithm for estimating link travel time,” *Intelligent Transport Systems*, vol. 7, no. 3 & 4, pp. 235-260, 2003.

[22] W. Pu, J. (Jane) Lin, and L. Long, “Real-Time Estimation of Urban Street Segment Travel Time Using Buses as Speed Probes,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2129, pp. 81-89, 2009.

[23] D. E. Lucas, P. B. Mirchandani, and N. Verma, “Online travel time estimation without vehicle identification,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1867, no. 1, pp. 193–201, 2004.

[24] S. M. Turner, T. J. Lomax, and H. S. Levinson, “Measuring and Estimating Congestion Using Travel Time-Based Procedures,” *Transportation Research Record*, vol. 1564, no. 1, pp. 11-19, Jan. 1996.

[25] T. Park and S. Lee, “A bayesian approach for estimating link travel time on urban arterial road network,” *Lecture notes in computer science*, pp. 1017–1025, 2004.

[26] R. L. Cheu, D.-h D.-H. Lee, and C. Xie, “An arterial speed estimation model fusing data from stationary and mobile sensors,” *ITSC 2001. 2001 IEEE Intelligent Transportation Systems. Proceedings (Cat. No.01TH8585)*, vol. pp, pp. 573-578, 2001.

[27] P. V. Palacharla and P. C. Nelson, “Application of fuzzy logic and neural networks for dynamic travel time estimation,” *International Transactions in Operational Research*, vol. 6, no. 1, pp. 145-160, Jan. 1999.

[28] S. Robinson and J. Polak, “Modeling Urban Link Travel Time with Inductive Loop Detector Data by Using the k-NN Method,” *Transportation Research Record*, vol. 1935, no. 1, pp. 47-56, Jan. 2005.

[29] N. Geroliminis and A. Skabardonis, “Prediction of Arrival Profiles and Queue Lengths Along Signalized Arterials by Using a Markov Decision Process,” *Transportation Research Record*, vol. 1934, no. 1, pp. 116-124, Jan. 2005.

[30] A. Skabardonis and N. Geroliminis, “Developement of Performance Measures on Signalized Arterials,” Berkeley, 2008.

[31] T. Tsekeris and A. Skabardonis, “On-line performance measurement models for urban arterial networks,” in *Transportation Research Board Annual Meeting*, 2004, vol. 935, p. 24.

[32] K. Kwong, R. Kavaler, R. Rajagopal, S. Networks, P. Varaiya, and C. Sciences, "A practical scheme for arterial travel time estimation based on vehicle re-identification using wireless sensors," in *TRB 2009 Annual Meeting (CD-ROM)*, 2009, no. 1750, p. 15.

[33] J. Luk, C. Karl, M. Su, and P. Bennett, "Real-time estimation of travel times on arterial roads in melbourne," in *22nd ARRB Conference – Research into Practice*, 2006, p. 14.

[34] J. Wahle, "A cellular automaton traffic flow model for online simulation of traffic," *Parallel Computing*, vol. 27, no. 5, pp. 719-735, Apr. 2001.

[35] H. H. Henry X. Liu, Wenteng Ma, Xinkai Wu, "Development of a Real-Time Arterial Performance Monitoring System Using Traffic Data Available from Existing Signal Systems," Minneapolis, 2009.

[36] INRIX Inc., "INRIX - Go Anywhere," 2011. [Online]. Available: http://www.inrix.com/. [Accessed: 01-Sep-2011].

[37] Clear Channel Radio, "Total Traffic Network - Navigating the Future," 2011. [Online]. Available: http://totaltraffic.com/. [Accessed: 01-Sep-2011].

[38] SpeedInfo, "SpeedInfo - Delivering Data that Knows the Flow," 2007. [Online]. Available: http://www.speedinfo.com/. [Accessed: 01-Sep-2011].

[39] D. Henclewood et al., “Real-Time Estimation of Arterial Travel Time and Operational Measures through Integration of Real-Time Fixed Sensor Data and Simulation,” Atlanta, GA, 2010.

[40] PTV, “VISSIM User Manual,” Karlsruhe, Germany, 2010.

[41] PTV, “VISSIM 5.10-06 COM Interface Manual,” Karlsruhe, Germany, 2009.

[42] Microsoft, “Microsoft Developer Network,” 2011. [Online]. Available: http://msdn.microsoft.com/en-us/.

[43] NGSIM, “The NGSIM Community Website,” 2011. [Online]. Available: http://ngsim-community.org/.

[44] C. Systematics, “NGSIM Peachtree Street (Atlanta) Data Analysis (12:45 p.m. to 1:00 p.m.),” Oakland, 2007.

[45] C. Systematics, “NGSIM Peachtree Street (Atlanta) Data Analysis (4:00 p.m. to 4:15 p.m.),” Oakland, 2007.

[46] D. Miller, “Developing a Procedure to Indentify Parameters for Calibratoin of a VISSIM Model,” Georgia Institute of Technology, 2009.

[47] D. M. Miller, D. Henclewood, M. Rodgers, and M. Hunter, “Parameter Selection Procedure for Signalized Arterial Simulation Calibration,” in *TRB 91st Annual Meeting Compendium of Papers CD-ROM*, 2012, vol. 7228.

[48] A. M. Law and W. D. Kelton, *Simulation Modeling and Analysis*, Second Edi. New York: McGraw-Hill, Inc., 1991, p. 759.

[49] B. Hellinga, "Requirements for the Calibration of Traffic Simulation Models," in *Canadian Society for Civil Engineering 1998 Annual Conference*, 1998, pp. 211-222.

[50] B. B. Park and J. D. Schneeberger, "Calibration and Validation Case Study of VISSIM Simulation Model for a Coordinated Actuated Signal System," *Transportation Research Record*, vol. 1856, pp. 185-192, 2003.

[51] B. (Brian) Park and J. Won, "Microscopic Simulation Model Calibration and Validation Handbook," Chapel Hill, NC, 2006.

[52] Transportation Research Board (TRB), "Highway Capacity Manual," Washinton DC, 2010.

[53] R. P. Roess, W. R. McShane, and E. S. Prassas, *Traffic Engineering*, Third Edit. 2003, p. 730.

[54] C. Z. Mooney and R. D. Duval, *Bootstrapping: A Nonparametric Approach to Statistical Inference*, Sage Unive. Newbury Park, CA: Sage: , 1993, p. 73.

[55] P. H. Kvam and B. Vidakovic, *Wiley Series in Probability and Statistics, in Nonparametric Statistics with Applications to Science and Engineering*. Hoboken, NJ, USA: John Wiley & Sons, Inc, 2007.

[56] R. Gentleman and R. Ihaka, “The R Project for Statistical Computing,” 1993. [Online]. Available: http://www.r-project.org/.

[57] T. Hothorn and K. Hornik, “Exact Distributions for Rank and Permutation Tests - The ‘exactRankTests’ Package,” 2011.

[58] G. von Mayr, *Die Gesetzmaßigkeit im Gesellschaftsleben*. Munchen: Oldenbourg, 1877, p. 354.

[59] A. May, *Traffic Flow Fundamentals*, 1st ed. Prentice Hall, 1997, p. 464.

[60] G.-L. Chang and X. Tao, “Estimation of Time-Dependent Turning Fractions at Signalized Intersections,” *Transportation Research Record*, vol. 1644, no. 1, pp. 142-149, Jan. 1998.

[61] X. Ban, R. Herring, P. Hao, and A. Bayen, “Delay pattern estimation for signalized intersections using sampled travel times,” in *2009 Transportation Research Board Annual Meeting,(Washington, DC)*, 2009, p. 21.

[62] W. Suh, D. Henclewood, A. Greenwood, A. Guin, R. Guensler, and M. Hunter, “MODELING PEDESTRIAN CROSSING USING MICRO- SCOPIC TRAFFIC SIMULATION,” in *TRB 2012 Annual Meeting (CD-ROM)*, 2012, vol. 6680, p. 15.